NEMATODES ADAPT USING YIN-YANG ISOFORMS OF A NURF SUBUNIT

A Dissertation

Presented to

The Academic Faculty

by

Wen Xu

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy in Biology in the

School of Biological Sciences

Georgia Institute of Technology

May 2019

NEMATODES ADAPT USING YIN-YANG ISOFORMS OF A NURF SUBUNIT

Approved by:

Dr. Patrick T. McGrath, Advisor
School of Biological Sciences
Georgia Institute of Technology

Dr. Greg Gibson
School of Biological Sciences
Georgia Institute of Technology

Dr. Jeffery T. Streelman
School of Biological Sciences
Georgia Institute of Technology

Dr. Annalise Paaby
School of Biological Sciences
Georgia Institute of Technology

Dr. David J. Katz
Department of Cell Biology
Emory University

Date Approved: [March 19, 2019]

ACKNOWLEDGEMENTS

First, I would like to give my sincere gratitude to my thesis advisor Patrick McGrath for his innumerous support in my research and personal life, for his patience, humor and wealth in knowledge. His guidance helped me to find my interest to fundamental genetics studies and inspired me to dedicate to my research. His support allowed me to learn more beyond biology and gave me the courage to explore fields that I was totally unfamiliar with. As I always say, Patrick is the best advisor in my heart.

Besides my advisor, I would like to thank the rest of my committee members: Professor Greg Gibson, Professor Todd Streelman, Professor Annalise Paaby, and Professor David Katz, for their comments to help me understand more about genetics, evolution, experimental techniques, scientific talks and more.

I would also like to thank past and present McGrath lab members: Edward Large, Harsh Patel, Nicole Baran, Mark Lowder, , Kathie Watkins, Richard Campbell, Lijiang Long, Yuehui Zhao, Weipeng Zhuo, Eric Anderson and many others, for the nights we found each other in the lab, for the projects we worked out together, for the supportive smiles and for the fun times we have had.

Outside of McGrath lab, I would like to thank my friends: Han-Ting Chou, Haozheng Tian, Tongtong Xu, Ying Sha, Dan Sun, Chinar Patil, Lu Wang, Jin Xu, and Cody Wang, for the good or bad time we spent, for the delicious food we ate, for the card games we played, and for the gossip we shared.

I would like to thank my collaborators at Georgia Tech, Emory University and Spanish National Cancer Research Center, Georgia Tech biology, biomedical engineering, the

chemical engineering community, Dalia Gulick, Naima Djeddar and Anton Bryksin from Core facilities, for their generous help.

Finally, I would like to thank my husband Tian Mi, my brother and my parents, for their continuous support and love. And, of course, my baby Lucas, for his most beautiful simile. Thank you for joining my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

ARL	Allelic Replacement Line
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DIL	Domain Isogenic Line
DSB	Double Strand Break
HR	Homologous Recombination
L4	Larval stage 4
lof	loss of function
MosSCI	Mos-1 induced Single Copy Insertion
NHEJ	Non-Homologous End Joining
NIL	Near Isogenic Line
PCA	Principal Component Analysis
QTL	Quantitative Trait Locus
SIL	STOP Isogenic Line
SNP	Single Nucleotide Polymorphism

SUMMARY

Convergent or parallel evolution is the repeated evolution of the same genotype in independent populations in response to similar environmental changes. A growing number of examples of parallel evolution are accumulating in the literature (e.g. cis-regulatory changes in the shavenbaby developmental regulator in Drosophila species result in dorsal cuticle hair loss [1], repeated selection on the Eda TNF ligand causes stickleback low-plated phenotype [2], and deletion of chemoreceptor genes contribute to the insensitivity to a specific pheromone in Caenorhabditis species [3]). In this dissertation, I discussed my studies of how Caenorhabditis elegans strains adapt to laboratory environments. I described how two C. elegans strains N2 and LSJ2, who share a common ancestor but have evolved independently in laboratory conditions have increased fitness in their respective environment. I showed that part of adaptation in the LSJ2 strain is caused by a 60 bp deletion in nurf-1 gene, a subunit of nucleosome remodeling factor NURF. Next, I described my finding about that adaptation of the N2 lineage is partially caused by a SNV (single nucleotide variation) in the 2nd intron of nurf-1. This work suggests that nurf-1 is a common target of evolution in response to laboratory growth. Finally, I described my work to understand why nurf-1 might be targeted, which I propose is due to the antagonistic function (here I refer as Yin-yang) of two major nurf-1 isoforms on the sexual fate during gametogenesis. My doctoral thesis study advances our understanding of how nucleosome remodeling factor may work and that isoform-level study of complex genes is feasible and necessary.

CHAPTER 1. INTRODUCTION

Does evolution follow predictable trajectories? Evolutionary biologist seeks to find out if determinism or contingency plays a major role in the evolutionary history of life. Stephen Jay Gould proposed the experiment about *replaying life's tape*. Will we see the world as it is today or a thoroughly different one? Gould argued that due to the contingency nature of evolution, if we rerun the tape over and over, we would see totally different outcomes [4]. However, performing such an experiment seemed to be impossible at Gould's time. Recent high-throughput experimental evolution studies have been able to *replay life's tape*, although in a smaller scale[5]. Often, similar genetic changes were found in these studies. For example, in high-throughput experimental evolution studies in yeast, 77 putatively adaptive mutants occurred in just six genes[6]; in *E. coli* heat shock response experimental evolution assay, 3 out of 5 independent duplication/deletion events occurred at the same genomic location[7].

Experimental evolutionary studies provided an opportunity for researchers to control certain environment and record the responses. However, there are also some limitations to this approach. Although genetic manipulation techniques are often used to test the phenotypic and fitness effect of particular variations, the limitation of time, population size, organism, and experimental conditions result in bias or constraint of results in these studies. Will the same gene or genetic location be repeatedly selected in nature?

With the current power of mapping and identifying genes harboring causative genetic variations, researchers often rediscover the same gene in different populations or species through phylogenetic studies and quantitative genetics studies[8]. Most early flowering

Arabidopsis ecotypes carry one of two different deletions in FRIGIDA locus that disrupt the open reading frame [9]; Studies of pigmentation differences have repeatedly mapped genetic variation in the *agouti* and the MC1R genes[10]–[13]; deletion of chemoreceptor genes contribute to the insensitivity to specific pheromone in *Caenorhabditis* species[3]. Finally, modifications to development or body plan have often identified *cis* variation in key transcription factors or master regulators of cell fate [1], [2], [14]. These studies show that similar phenotypic change can result from common genetic targets.

Adaptation is a gradual process. A substantially good resource to study adaptation is critical for identification of adaptive variants. For evolutionary genetics study of model organisms, wildtype animals are collected and assayed through quantitative trait locus (QTL) mapping to find regions that are responsible for phenotypic differences. This often results in the identification of one or multiple regions (the QTLs) carrying various variations. Due to the

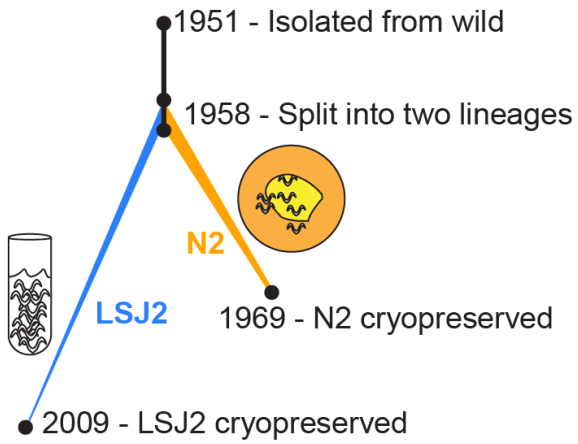

Figure 1.1. **Laboratory adaptation of two *Caenorhabditis elegans* strains N2 and LSJ2.** History of the C. elegans strains N2 and LSJ2 following isolation from the wild (Bristol, England). LSJ2 was grown in liquid axenic culture whereas N2 was propagated on agar plates using bacteria as food source. 282 variations were fixed between these two strains, 188 in LSJ2 lineage and 94 in N2 lineage.

complex genetic background of natural isolates, one QTL can contain hundreds to thousands of variations, which makes identifying the quantitative trait nucleotides(QTNs) difficult. In McGrath lab, we utilize two *C. elegans* strains N2 and LSJ2, both are descendants of a worm strain isolated in 1951 from Bristol, England. The original culture was separated into two in 1957 and the N2 lineage has been maintained on agar plates with a bacterial food source (*E. coli* OP50) until it was frozen in 1969 (**Figure 1.1**). The LSJ2 lineage was mistakenly thought to be *Caenorhabditis briggsae* – another *Caenorhabdiditis* specie that is closely related to *C. elegans* – and was cultured in axenic culture consisting of soy-peptone extract supplemented with beef liver extract as food source from the time of separation until it was frozen in 2009 as LSJ2. We found that 94 new mutations were fixed in the N2 lineage and 188 new mutations were fixed in the LSJ2 lineage [3]. This genetic diversity is almost three orders of magnitude lower than the genetic diversity between wild strains of *C. elegans*[15]–[17], and four orders of magnitude lower than the genetic diversity between two humans[18], making identification of causative mutations through QTL mapping feasible.

My advisor previously found a hotspot locus - a deletion disrupts chemoreceptor genes that are responsible for pheromone sensing - using this two laboratory adapted *Caenorhabditis elegans* strains, while the similar event was also found in *C. briggsae* [3]. Recently, my lab mates Edward Large and Yuehui Zhao discovered a QTL on the right arm of chromosome II by measuring egg-laying behavior and fitness of Recombinant Inbred Lines (RILs) of N2 and LSJ2 **(Figure 1.2A)**[19]. By further analysis, they refined the causative locus to a region containing 11 genetic variations within this region: three in intergenic regions, six

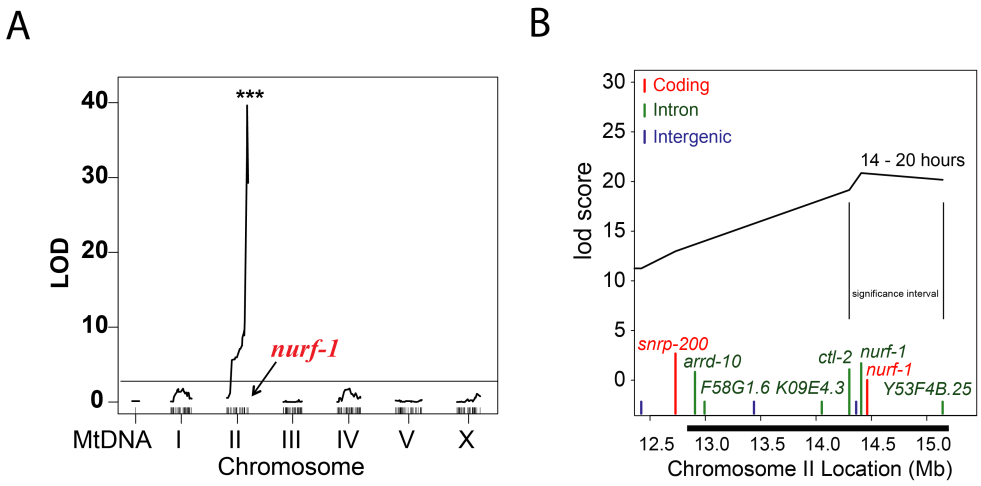

Figure 1.2 – The major effect QTL on chromosome II. QTL mapping of egg-laying behavior and fitness of N2 and LSJ2 mapped to the right arm of chromosome II. **A)** QTL mapping of egg-laying rate found a QTL locus contains *nurf-1* gene. **B)** 11 genetic variations under the chromosome II QTL. Two variations were in the coding region, six in intron and three in intergenic regions.

in intronic regions and two in coding regions (**Figure 1.2B**). To further understanding the causative variations, precise genetic manipulations are required.

Recent years, the genome editing system CRISPR/Cas9 has become a popular tool due to its convenience and efficiency. Previous known as the bacterial immune system to defense virus infection, the Cas9 nuclease utilize Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) sequence as a guide to creating double-strand break (DSB) at specific DNA locus that is integrated into these repeats. This guide sequence is restricted to be in front of a short 3-5bp sequence (NGG or NGAG) called protospacer adjacent motif (PAM), which limits the guide sequences design. In *C. elegans*, the genome content is 66% AT[20] thus making it more difficult to find a PAM site. Moreover, the guiding efficiency is not completely understood and the guide sequence lead to DSB should not be too far

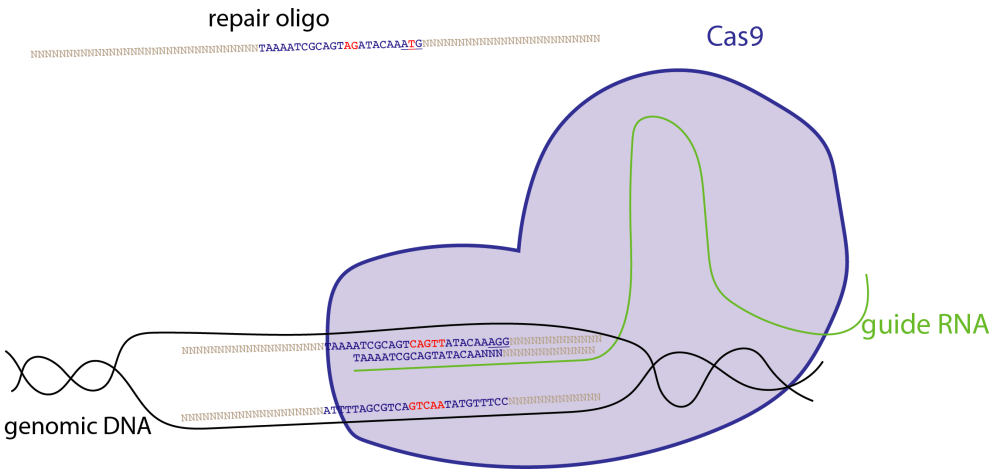

Figure 1.3 – CRISPR/Cas9 genome editing. CRISPR/Cas9 system uses a single guide RNA sequence (in blue) with an NGG PAM site (underlined sequence) that leads Cas9 nuclease to create a DSB three base pairs before the PAM. When repair oligo containing target mutation (in red) is provided, the DSB can integrate the target mutation into the genome through homologous recombination.

away from the targeted mutation site. After creating a DSB, the cell repair machinery can either recover through non-homologous end joining (NHEJ) or homologous recombination (HR) if a repair oligo with homologous arms is provided (**Figure 1.3**). By using CRISPR/Cas9 genome editing system with repair oligo, we will be able to reconstitute most variations in all genetic background.

In chapter two of this dissertation, I take advantage of the well-developed co-conversion CRISPR/Cas9 system to study the effect of a *nurf-1* deletion variation evolved from LSJ2 lineage. It is one of two coding variations in chromosome II QTL.

Chapter three investigate the fitness effect of another variation falls on *nurf-1* gene, which is an intron SNP evolved from N2 lineage. This intron SNP is hard to obtain through CRISPR/Cas9 genome editing since it is in a poly(A) and poly(T) region. Two-step CRISPR/Cas9 was applied to change the N2 intron allele to LSJ2 version.

Finally, chapter four is a detailed investigation of the function of *nurf-1* gene at isoforms level by utilizing the CRISPR/Cas9 system together with classical genetics methods.

CHAPTER 2. Adaptation of LSJ2 Lineage is Partially Caused By a 60bp Deletion in *nurf-1* gene

2.1 Introduction

The N2 strain is widely used in *C. elegans* labs as a reference wild type strain. But it has acquired previously described genetic variants in the *npr-1* and *glb-5* genes that affect a number of physiological traits[21]–[24]. To avoid studying these previously described laboratory adaptations, I utilized a strain, CX12311, which contains ancestral alleles of these two genes backcrossed to an N2 background. This strain will be referred to as N2* for simplification. Despite this low level of genetic diversity of N2* and LSJ2 (in total 282 fixed variations), a large number of phenotypic differences distinguish these two strains (**Table 2.1**). Previous QTL mapping discovered a chromosome II QTL for fitness and egg laying differences. Further examination by making a *nurf-1* Near Isogenic Line (NIL_{nurf-1}) refined this region to ~2Mb containing 11 genetic variations (**Figure 1.2B**), within which a *nurf-1* 60bp deletion resides. Only two coding variations were found in this chromosome II QTL and *nurf-1* 60bp deletion is within the significance region. I became interested in this 60bp deletion that deletes the 3' end of *nurf-1* gene, resulting in loss of 18 amino acids at the end of C-terminal coding region and 8bp of the 3'-UTR. My first hypothesis was that the deletion variation is responsible for the fitness difference between N2* and LSJ2.

Table 2.1. Phenotypic differences between CX12311(N2*) and LSJ2.

	CX12311(N2*)	LSJ2
growth rate	faster	slower
reproduction	faster	slower
pheromone-induced entry to dauer	yes	resistant
life span	shorter	longer
drug resistance	lower	higher
individual survival vs. reproduction	reproduction	individual survival

2.2 Results and Discussion

2.2.1 A 60bp deletion in nurf-1 *gene under the major-effect QTL partially explains the fitness differences of N2 and LSJ2*

In order to determine the LSJ2 *nurf-1* 60bp deletion's effect on fitness, I created a *nurf-1* allelic replacement line (ARL_{del} - PTM88) in N2* background (**Figure 2.1A**). This ARL_{del} strain has the exact same deletion in *nurf-1* gene. With the help from Yuehui Zhao, this strain was sequenced using next-generation whole genome sequencing technique to test if

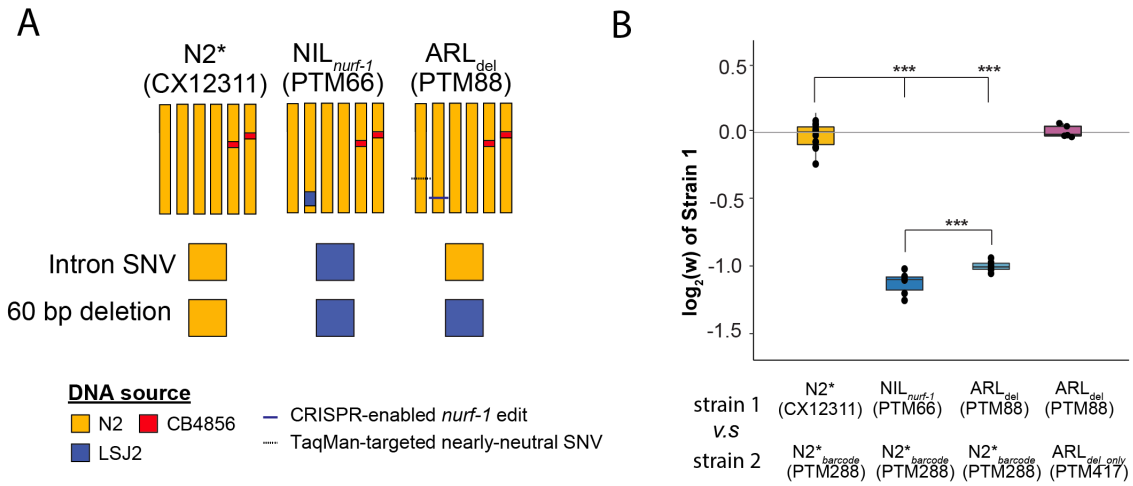

Figure 2.1. ARL*del* explains part of the fitness decrease of LSJ2 on agar plates. A) Animation illustration of NIL*nurf-1* and ARL*del* genetics. *npr-1* and *glb-5* ancestral alleles from CB4856 are labelled in red. **B)** Fitness coefficient of competition assay for NIL*nurf-1* ARL*del* and ARL*del_only*. y-axis indicates the relative fitness coefficient of strain 1 versus strain 2. Three stars indicate significant differences ($p<0.01$).

there exist other variations that were maintained during strain construction. A *spe-9* SNP (kah132) was discovered in this strain. It was an A to G mutation in the last fourth exon of *spe-9* gene and result in an amino acid change from asparagine to aspartic acid. This variation came from the propagation of ARL*del* strain before cryopreserved. To take *spe-9(kah132)*'s effect on fitness into consideration, I created a control strain ARL*del_only* (PTM417) by backcrossing ARL*del* to N2*. ARL*del_only* only contains the *nurf-1* 60bp deletion and was used as a control strain for the competition experiment. The average $\log_2$ relative fitness coefficient of ARL*del_only* against ARL*del* is -0.00685 (**Figure 2.2B**), suggesting no or little effect of *spe-9* SNP.

A competition assay was then performed to compare relative fitness of different strains with the help from Yuehui Zhao. Average $\log_2$ relative fitness coefficient of NIL*nurf-1* was -1.1288 while for ARL*del* was -1.0046. Results suggest ARL*del* was less fit than N2*, which

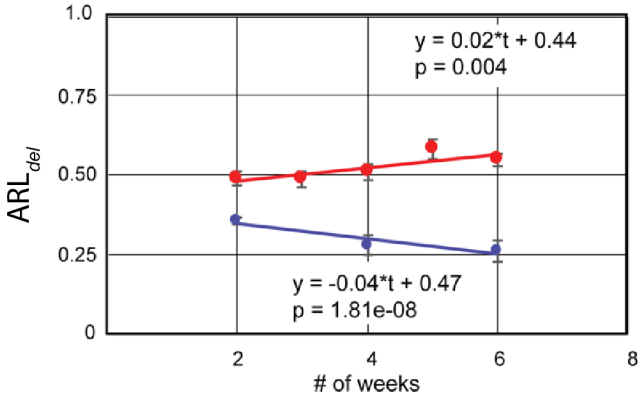

Figure 2.2. 60bp deletion in *nurf-1* is advantageous in liquid culture but disadvantageous on an agar plate. The x-axis is the time point when worms were sampled for genotype quantification, y-axis indicates the allele frequency of ARL_{del} for each time point. Competition assay performed on agar plates were labeled in red and in liquid culture is labeled in blue.

explained about 89% of the fitness disadvantage of $NIL_{nurf\text{-}1}$. This suggests the fitness disadvantage of LSJ2 on agar plates is partially but also mostly caused by the 60bp deletion in *nurf-1* gene.

I next asked if the *nurf-1* 60bp deletion had a different effect in environments that N2 and LSJ2 evolved in. Yuehui performed competition assay both on agar plates and in liquid culture[19]. Results show that this deletion variation is advantageous in liquid culture while disadvantageous on agar plates (**Figure 2.2**).

2.2.2 The nurf-1 *deletion cause animals to grow slower and reproduce later*

To further investigate how *nurf-1* 60bp deletion affect fitness, I tested two fitness proximal traits for ARL_{del}. The egg laying rates of ARL_{del} were significantly different from N2* on the third, fourth and fifth time-points in the direction that's consistent with $NIL_{nurf\text{-}1}$, but it is still distinguishable from $NIL_{nurf\text{-}1}$ (**Figure 2.3A**). Next, I tested the growth rate of

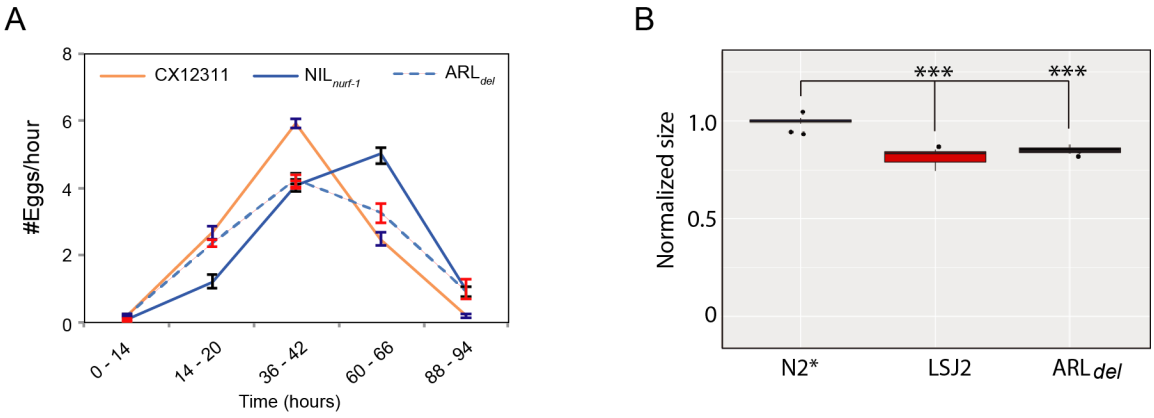

Figure 2.3. 60bp deletion in *nurf-1* cause animals to reproduce later and grow slower. **A)** Egg laying rate at representative timepoints indicates later reproduction timing of NIL_{nurf-1} and ARL_{del}. the y-axis is the egg laying rate at specific time points. **B)** The growth rate of N2*, LSJ2 and ARL_{del}. the y-axis shows the average area of each strain measured through video tracking relative to the average area of N2*. *** suggest a significant difference ($p<0.01$).

animals by synchronizing animals at hatching followed by video recordings of the animals at 72 hours. Animals increase their volume 100-fold during the course of experiments. These videos were analyzed by Matlab codes to identify the pixel area of the animals normalized to the average size of N2* strain. LSJ2 and ARL_{del} animals grew at a significantly slower rate than N2* (**Figure 2.3B**). Results from two fitness proximal traits suggest reproduction and growth regulation are two of the ways how *nurf-1* affect fitness.

2.3 Methods

2.3.1 Strains

Strains used in this chapter are: N2, LSJ2, CX12311 (N2*) *kyIR1(V, CB4856>N2); qgIR1(X, CB4856>N2)* , PTM88 (ARL_{del}) *nurf-1(II, kah3); spe-9 (I, kah132); kyIR1(V, CB4856>N2); qgIR1(X, CB4856>N2),* PTM93 (ARL_{del}) *nurf-1(II, kah5);*

kyIR1(V, CB4856>N2); qgIR1(X, CB4856>N2), PTM66 (NIL$_{nurf\text{-}1}$) *kyIR87(II, LSJ2>N2); kyIR1(V, CB4856>N2); qgIR1(X, CB4856>N2),* PTM288 (N2*_barcode) *dpy10(kah83); kyIR1 (V, CB4856>N2); qgIR1 (X, CB4856>N2) X,* PTM417 (ARL$_{del_only}$) *nurf-1(kah3); kyIR1(V, CB4856>N2); qgIR1(X, CB4856>N2)*

2.3.2 CRISPR ARL$_{del}$ strain construction

I generated the ARL$_{60bp_del}$ strain following the published co-conversion CRISPR method to simultaneously edit the *dpy-10* gene (as a visual marker) along with *nurf-1* using single-stranded oligonucleotides as repair templates[25]. All sgRNAs were cloned into a subclone of pDD163 containing the U6 promoter to drive sgRNAs in the germline[26]. For the *dpy-10* gene, I used the previously published sgRNA and repair oligo. For the *nurf-1* gene, I designed a sgRNA to target the 5'-TTCGGATCAGCTGTTGCCAC(TGG)-3' protospacer/PAM site found in the LSJ2 60bp deletion. I used single-stranded oligonucleotide 5'-TCTATCAGAAAGCGTGTCCAGTCGGAAAGCCAGCGAACTGT CGACTCGTTGGATATCGATTCCTCTTGTTTTTTTATGTTTTTCGTAGTCACACA GTGACTTTTCACTTGTTACGTTGACAATGT -3' as a repair construct. To drive Cas9 in the germline, I subcloned P*eft-3*::Cas9 from pDD162 into a separate vector. I injected 50 ng/ul P*eft-3*::Cas9, 25 ng/ul *dpy-10* sgRNA, 500 nM dpy-10(cn64) repair oligo, 25 ng/ul *nurf-1* sgRNA, and 500 nM LSJ2 *nurf-1* repair oligonucleotide into CX12311 animals. I genotyped 113 F1 roller animals by PCR using the primers 5'- ACATTATACG AAGTTATGTCGTCAAACTTTGCATTTG-3' and 5'-CATCTTCATAATTCCAACGG AAACCAAG-3' followed by digestion with PvuII (a site which is removed by the 60 bp LSJ2 deletion). I identified a single PvuII resistant band, however, Sanger sequencing showed that this F1 animal contained a 10 bp deletion of 5'-AGCTGTTGCC-3' replaced

by 5'-GA-3', indicating that this lesion resulted from an NHEJ event that disrupted the PvuII site as opposed to our targeted HR from the *nurf-1* oligo.

I hypothesized that due to the 60 bp deletion, the flanking regions on the single-stranded oligonucleotide were not long enough to initiate homologous repair. I next generated a strain, PTM91, containing an extrachromosomal array of the LSJ2 *nurf-1* 3' region by injecting a PCR product generated using the 5'-GCAATTTGTGAACGACGTGA-3' and 5'-CCGGTCTCGACACAATTTTT-3' primers along with a P*elt-2*::GFP co-injection marker into CX12311 animals. I injected the co-conversion injection mix described above into these animals and again singled 80 F1 roller or roller/dumpy animals and genotyped them as above. I identified two PvuII-resistant bands, which Sanger sequencing showed was due to the presence of the LSJ2 60 bp deletion in both strains. These two strains were dumpy rollers (indicating a conversion event along with a deletion event in the *dpy-10* locus). I mated these two strains to CX12311 males to separate the *nurf-1* deletions from the *dpy-10* mutations.

2.3.3 Competition assay for measuring fitness

Competition assays were mostly performed on NGM plates except that CX12311 vs. ARL_{del} (PTM88) assay was also performed in liquid HS-YE-HLE stock media. The experimental design is shown in **Table 2.2**.

In the competition assay on NGM plates seeded with OP50 bacteria, six experimental replicates were performed. At the beginning of each assay, ten L4 animals of both strain 1 and strain 2 were placed on 9cm NGM plates and kept at 20°C until starve. At this point, animals were washed by 1ml M9 to 1.5ml Eppendorf tubes and let precipitate for one

Table 2.2. Experimental design of competition assay for *nurf-1* 60bp deletion related strains.

Strain 1	Strain 2	Taqman Probe target
N2* (CX12311)	N2*$_{barcode}$ (PTM288)	*dpy-10(kah83)*
NIL$_{nurf-1}$ (PTM66)	N2*$_{barcode}$ (PTM288)	*dpy-10(kah83)*
ARL$_{del}$ (PTM88)	N2*$_{barcode}$ (PTM288)	*dpy-10(kah83)*
ARL$_{del}$ (PTM88)	ARL$_{del_only}$ (PTM417)	*spe-9 (kah132)*
ARL$_{del}$ (PTM88)	N2* (CX12311)	*dpy-10(kah83)*

minute. 50ul of upper layer M9 containing only L1s was then transferred new NGM plates. Populations were continuously cultured in this way for 7 generations or 5 weeks. Genomic DNA was isolated from the 1st, 3rd, 5th and 7th transfer/week. The genomic DNA was isolated by using a Qiagen Gentra Puregene Zymo DNA isolation kit (D4071).

In the competition assay in liquid HS-YE-HLE stock media, six experimental replicates were performed. Each replicate was cultured in a 10 mL HS-YE-HLE stock media in a cell culture flask. At the beginning of this assay, the worms were synchronized by alkaline-bleach. When the animals were hatched, 100 L1 animals from each line were transferred into 10 mL of HS-YE-HLE media and cultured at 20°C in a vertical shaker. After two weeks, 1 mL of depleted culture (containing ~2000 animals) were transferred to a new cell culture flask container. Then, the populations were continuously cultured for another four

weeks. The populations were transferred every two weeks and the populations' genomic DNA was isolated at the 2nd, 3rd, 4th, 5th and 6th-week time point by the same method described above.

The proportion of strain 1 population density in a competition assay was measured using Taqman analysis in Biorad QX200 digital PCR system to quantify the frequency of probe target described in **Table 2.2**. TaqMan probes were designed using standard software from Applied Biosystems. Genomic DNA from each time point was digested with SacI enzyme and purified with Zymo DNA Clean & Concentrator Kit (cas. nos. D4004). The concentration of fragmented genomic DNA was adjusted to 2 ng/uL by Qubit assay (cas. nos. Q32851). Digital PCR was performed followed the standard method provided by Biorad with the absolute quantification method. The fitness coefficient was then calculated using the following equations.

$$P(A)_t = \frac{\text{No. of Allele A}}{\text{No. of Allele A} + \text{ No. of Allele a}}$$

$$P(A)_t = \frac{P(A)_0 W_{AA}^t}{P(A)_0 W_{AA}^t + (1 - P(A)_0) W_{aa}^t}$$

$$\log\left(\frac{\frac{P(A)_0}{P(A)_t} - P(A)_0}{1 - P(A)_0}\right) = \left(\log\left(\frac{W_{aa}}{W_{AA}}\right)\right) t$$

2.3.4 Egg laying rate analysis

All egg-laying assays were carried out at 20°C using standard 3cm NGM plates seeded with the OP50 strain of *Escherichia coli*. OP50 were prepared freshly by streaking a glycerol stock of OP50 on an LB plate and let grow at 37°C overnight. A single colony

was then picked to 5ml fresh LB and cultured overnight in a shaking incubator at 200rpm. 1ml of the overnight culture was used to inoculate 200 ml of LB for 4–6 hours of growth at 37°C with shaking. The 200ml OP50 culture was concentrated via centrifugation to an OD_{600} of 2.0 and this culture was used for seeding experimental plates with 50 µl aliquots. All experimental plates were prepared the week of the assay and left at 22.5°C 18–24 hrs following seeding. Plates were then placed at 4°C until the day of the assay and warmed to 20°C for 12 hours before each time point.

Nematodes were cultured at least three generations prior to the beginning of the assay. Six fourth larval stage (L4) nematodes were transferred to the first 50µl experimental plate. Five time points (0–14, 14–20, 36–42, 60-66, 88-94 hours) were measured for the number of eggs laid, and eggs laid per hour was calculated by dividing the time range and the number of animals left on each plate at each time point.

2.3.5 Growth rate analysis

Animals were synchronized by allowing 10-20 adults to lay eggs on an NGM plate seeded with OP50 bacteria for two hours and let grow at 20°C. Three replicates were created for each strain. At 72 hours, animals were recorded for one minute using a Video tracking camera. Matlab code was used to track worms and calculate the average size of each worm. The average size of animals from each plate was then normalized to the average size of the three N2 plates.

CHAPTER 3. A *nurf-1* Intron SNP Evolved From N2 Lineage Also Affect Fitness

3.1 Introduction

Growing evidence suggests that evolution can follow predictable genetic paths in response to specific environmental shifts. Laboratory conditions, and their associated selective pressures represent new environments faced by model organisms after their isolation from the wild. I showed in Chapter 2 that the LSJ2 lineage of Caenorhabditis elegans adapted to laboratory cultures in part through a 60bp deletion in the *nurf-1* gene. While I also showed that the 60bp deletion may not be the only variation that affects the fitness of N2 and LSJ2 due to the incomplete explanation of ARL_{del} to NIL_{nurf-1} on fitness and reproduction.

While reviewing the variations under the chromosome II QTL, I found another intron variation evolved from N2 lineage also fall in the *nurf-1* gene region. Genetic drift may play a role, but two out of the 282 fixed variations fall on a single gene that has fitness effect may not be a coincidence. Furthermore, this *nurf-1* intron SNP is derived from N2 lineage which inspired me to find out if this SNP was selected to increase fitness advantage for animals on agar plates. My main hypothesis for this chapter is that this intron SNP can also affect fitness.

3.2 Results and Discussion

3.2.1 Transcriptome analysis suggests another nurf-1 related variation in NIL_{nurf-1} background is causative

To investigated how ARL_{del} deletion and NIL_{nurf-1} affect transcriptome and find clues for other causative variations, Lijiang Long helped me sequenced the RNA from N2*, LSJ2, NIL_{nurf-1}, ARL_{del} young adult animals. From the transcriptome analysis, there exist hundreds to thousands of genes that show expression changes (**Table 3.1**).

Table 3.1. Number of up-, down- and total number of differentially expressed features for each comparison.

TEST VS REF	# DOWN	# UP	# TOTAL
ARL_{del} VS N2*	928	1373	2301
LSJ2 VS N2*	464	165	629
NIL_{nurf-1}VS N2*	737	413	1150
LSJ2 VS ARL_{del}	873	369	1242
NIL_{nurf-1}VS ARL_{del}	1356	321	1677
NIL_{nurf-1} VS LSJ2	78	36	114

A multidimensional PCA plot (**Figure 3.1A**) suggest LSJ2 and NIL_{nurf-1} affect transcriptome in similar ways and only 114 genes are differentially expressed (**Table 3.1**). From **Table 3.1**, we know that ARL_{del} and N2* have 2301 differentially expressed genes, while LSJ2 and N2* only have 629. This is surprising to us since we concluded in Chapter 2 that the 60bp deletion is causative for the difference in fitness of LSJ2 and N2*. After further inspection of the differentially expressed genes, we plotted the log2 fold change for

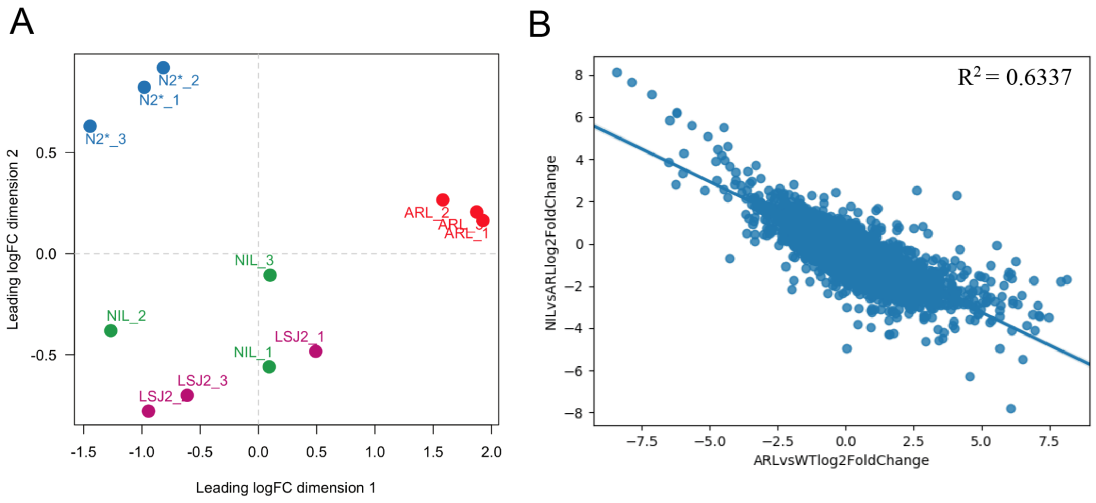

Figure 3.1. RNAseq analysis suggests another *nurf-1* related variation affecting NIL*nurf-1*. A) Multidimensional scaling plot of RNAseq results for N2*, LSJ2, NIL*nurf-* and ARL*del*. x- and the y-axis represents the two principal components that best separate the four groups. B) x-axis indicates the $\log_2$ Fold Change of ARL*del* vs. N2* and the y-axis the $\log_2$ Fold Change of NIL*nurf-1* vs. ARL*del*. Each dot represents a single gene, all genes detected in RNAseq experiment were plotted. A linear regression line was fitted with 0.6337 R^2 scores.

different comparison groups (**Figure 3.1B**). Interestingly, we found the genes upregulated in NIL*nurf-1* vs. ARL*del* group are downregulated in ARL*del* vs. N2* comparison and vice versa (R^2 = 0.6337) (**Figure 3.1B**). This result suggests another variation within the 11 variations found on the chromosome II QTL (**Figure 1.2B**) affect the same set of genes as the *nurf-1* 60bp deletion, but in the opposite direction. This preliminary result supported my hypothesis that the *nurf-1* intron SNP may not be a by-product of random drift.

3.2.2 The effect of nurf-1 intron SNP remains elusive after reciprocal heterozygosity test

To determine the effect of *nurf-1* intron SNP, I first tried to reverse the intron SNP in N2 back to its ancestral state using CRISPR/Cas9 genome editing. The intron SNP is in the 2nd

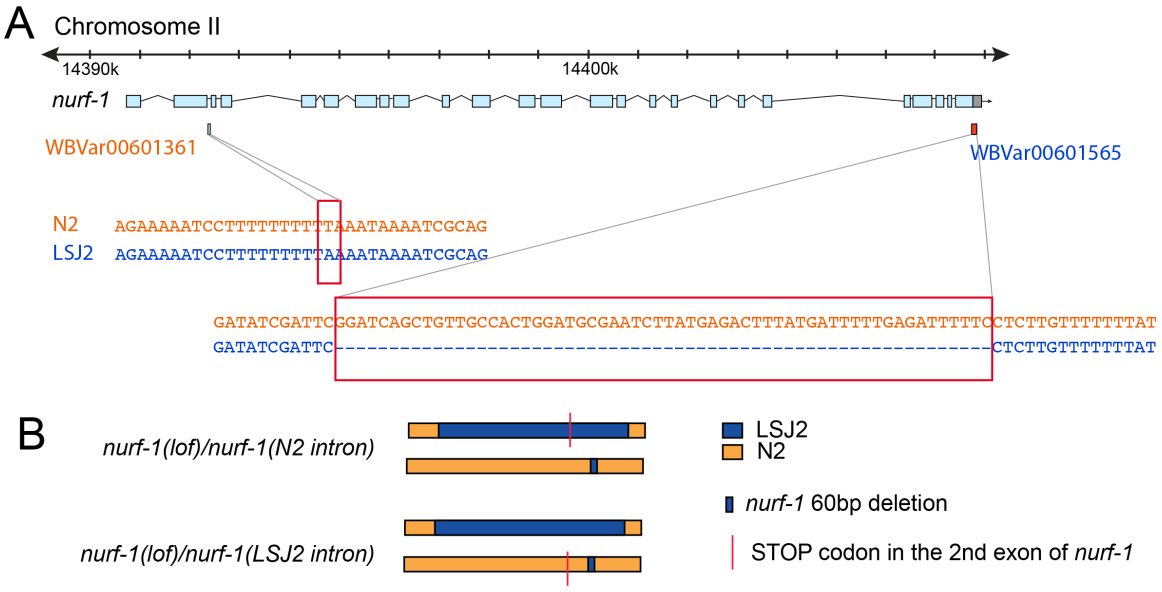

Figure 3.2 *nurf-1* intron variation and its reciprocal heterozygosity test design. A) WBVar00601361 is the intron SNP (ancestral A to T) in the 2nd intron that derived from N2 lineage and WBVar00601565 is the 60bp deletion derived from LSJ2 lineage that deletes GGATCAGCTGTTGCCACTGGATGCGAATCTTATGAGACTTTATGATTTT TGAGATTTTTC. **B)** Reciprocal heterozygosity test design for *nurf-1* intron SNP. CRISPR/Cas9 induced STOP codon in the 2nd exon of *nurf-1* is designed to mutate *nurf-1* gene the intron SNP can possibly affect, it then designated as *nurf-1(lof)* loss of function mutation. For *nurf-1(lof)/nurf-1(N2 intron)* strain, one copy of the chromosome carries a STOP codon in NIL_{nurf-1} background and the other chromosome is from ARL_{del}, which result in one functional copy of *nurf-1* from ARL_{del}. Similarly, *nurf-1(lof)/nurf-1(LSJ2 intron)* strain will carry a functional copy of *nurf-1* from NIL_{nurf-1}.

the intron of *nurf-1* and is an A to T change in a polyA and polyT region, which makes finding a good guide RNA with PAM site to be difficult (**Figure 3.2A**). I next tried a classical genetics method – reciprocal heterozygosity test, the experimental design is shown in **Figure 3.2B**.

The reciprocal heterozygosity test based on the assumption that *nurf-1* intron SNP only affect *nurf-1* gene and a STOP codon in the 2nd exon is a loss of function allele for *nurf-1*.

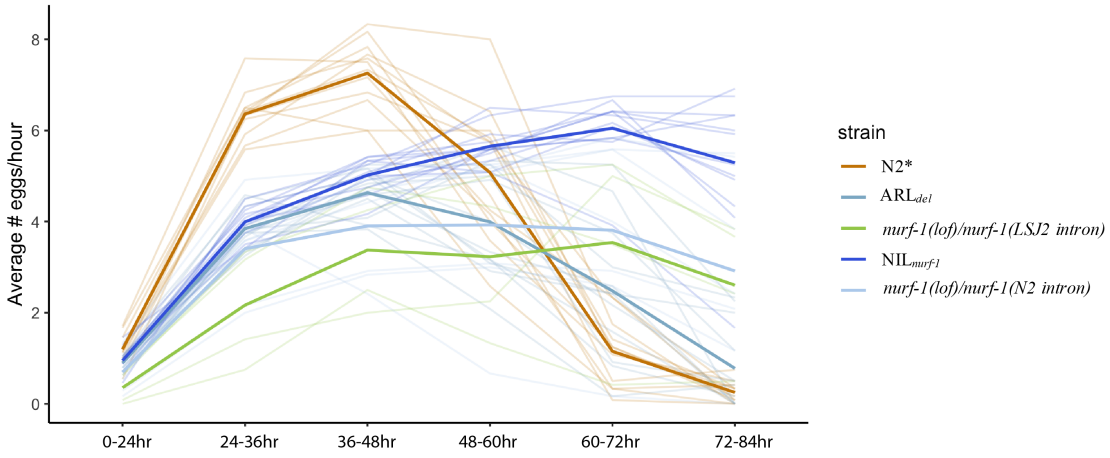

Figure 3.3. Egg laying rate analysis of reciprocal heterozygosity strains. N2*, NIL, ARL, *nurf-1(lof)/nurf-1(LSJ2 intron)* and *nurf-1(lof)/nurf-1(N2 intron)* strains were tested at 6 different time points. the y-axis shows the average number of eggs laid at specific time points. Significant test results are shown in **Table 3.2.**

Also, dosage effect is not considered in this design. My hypothesis is that *nurf-1(lof)/nurf-1(LSJ2 intron)* strain will have similar egg laying pattern (slower reproduction) as NIL$_{nurf-1}$ since the functional copy of *nurf-1* is the same – has LSJ2 *nurf-1* intron and 60bp deletion. Similarly, I expect *nurf-1(lof)/nurf-1(N2 intron)* will phenocopy ARL$_{del}$. But the egg laying rate results was out of expectation. *nurf-1(lof)/nurf-1(LSJ2 intron)* and *nurf-1(lof)/nurf-1(N2 intron)* strains actually didn't phenotype either NIL$_{nurf-1}$ or ARL$_{del}$, but are intermediate of the two (**Figure 3.3** and **Table 3.2**). This result does not convince me to make a clear conclusion on the effect of *nurf-1* intron SNP. This test was based on multiple assumptions that were hard to test. So I decided to focus on creating an Allelic Replacement Line for the intron SNP – ARL$_{intron.}$

Table 3.2. T-test p-value for reciprocal heterozygosity test egg laying rate at different timepoints. Comparisons that were not significantly different at p<0.1 level were labeled in grey.

t-test	**NIL*nurf-1* VS. *nurf-1(lof)/nurf-1(LSJ2 intron)***	**NIL*nurf-1* VS. *nurf-1(lof)/nurf-1(N2 intron)***	**ARL*del* VS. *nurf-1(lof)/nurf-1(LSJ2 intron)***	**ARL*del* VS. *nurf-1(lof)/nurf-1(N2 intron)***
0-24 hours	0.00243825	0.10891305	0.00136758	0.15978758
24-36 hours	0.00027578	0.0758301	0.00104643	0.19870239
36-48 hours	0.0011579	0.00570541	0.01116901	0.06560921
48-60 hours	0.0003139	0.00431213	0.33312081	0.92480128
60-72 hours	0.00318533	0.00322011	0.29798661	0.12733634
72-84 hours	0.00716936	0.00962852	0.01115418	0.00638976

3.2.3 ARL$_{intron}$ competition assay suggest nurf-1 intron SNP could affect fitness

The schema of ARL*intron* design is shown in **Figure 3.4A**. It is a two-step process: first, two sgRNA targeting the end of 2nd exon and beginning of 3rd exon were used with a repair oligo integrated with an efficient *dpy-10* sgRNA site to make an intermediate strain that carries the efficient targeting site; second, another CRISPR editing was applied by using *dpy-10* sgRNA with LSJ2 PCR product as repair template. Finally, I could get the intron SNP changed back to its ancestral state (LSJ2 version).

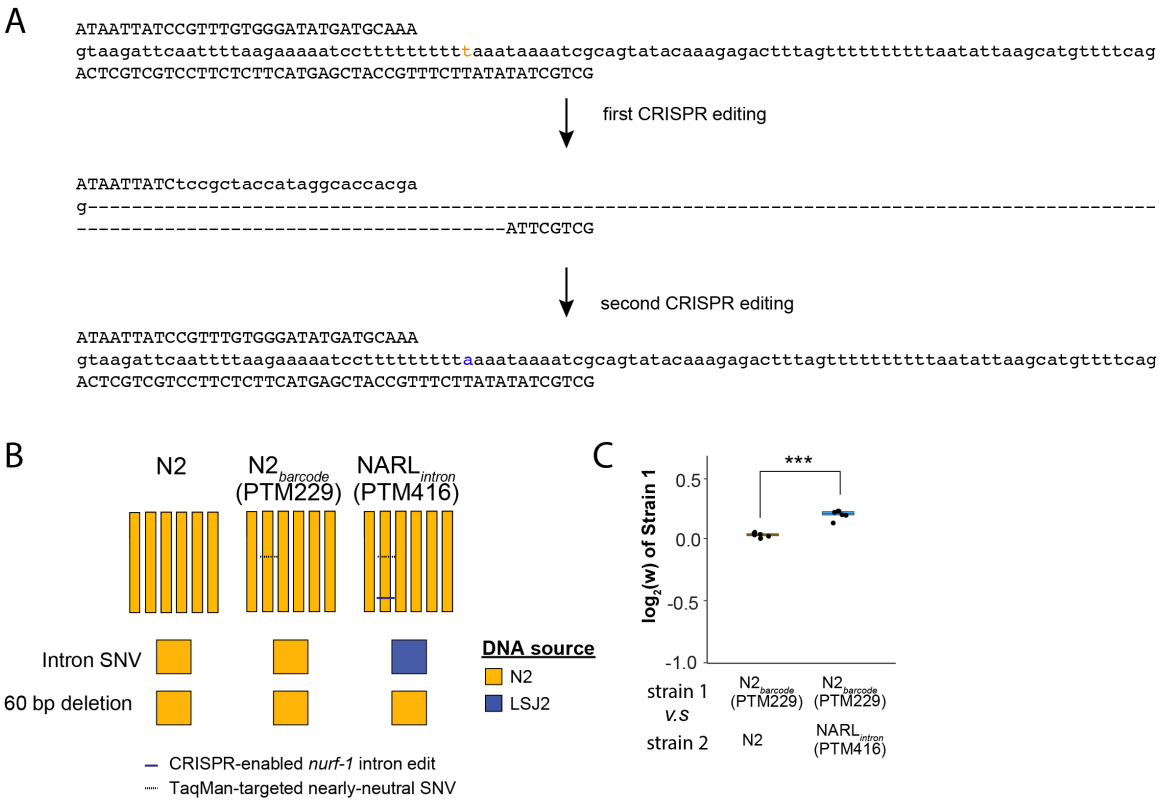

Figure 3.4. Two-step CRISPR design and competition assay for NARL*intron*. A) Two-step CRISPR/Cas9 editing was performed to generate NARL*intron*. Capitalized characters indicate exon sequences, lower case characters indicate intron sequence. WBVar00601361 SNP change is colour coded: orange character indicate N2 allele, blue character indicate LSJ2 allele. **B)** Animation illustration of NARL*intron* genetics. **C)** N2 had significantly higher fitness than NARLintron. y-axis indicates the relative fitness coefficient of strain 1 versus strain 2. Three stars indicate significant differences ($p<0.01$).

Since this *nurf-1* intron SNP was derived from N2 lineage, I want to test its effect in N2 background. So after getting the ARL*intron* strain, I backcrossed it to N2, who does not contain the *npr-1* and *glb-5* ancestral alleles, and named it as NARL*intron.* I then performed

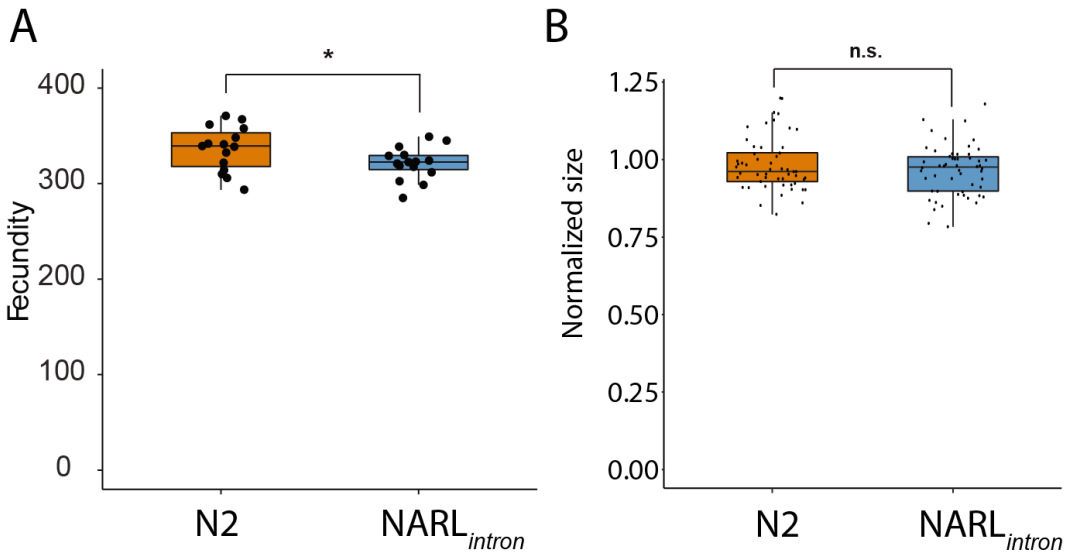

Figure 3.5. Fitness proximal traits of NARL*intron* and N2. A) Fecundity of N2 and NARL*intron*. the x-axis indicates the total number of eggs laid during 96 hours after L4 stage. One star indicates significant differences at p-value<0.1 level. **B)** Growth analysis of N2 and NARL*intron.*Animals' size was normalized to the average size of N2. n.s. indicates not significant at p-value<0.1 level.

competition assay for NARL*intron* against N2 and found NARL*intron* strain was less fit than N2 on agar plates with $\log_2$ relative fitness equals -0.158557 (**Figure 3.4B**). This result suggests that the *nurf-1* intron SNP evolved in N2 lineage is advantageous to fitness.

3.2.4 nurf-1 intron SNP regulate fitness through regulating reproduction

To understand how *nurf-1* intron affect fitness, I measured its fecundity and growth. NARL*intron* worms lay a similar number of eggs with N2 but with a very small difference (p-value <0.1) (**Figure 3.5A**). Results suggest that the intron SNP have an effect on reproduction but not growth (**Figure 3.5B**).

3.2.5 nurf-1 intron SNP affects a large number of genes during L4 to adult transition but not at the adult stage.

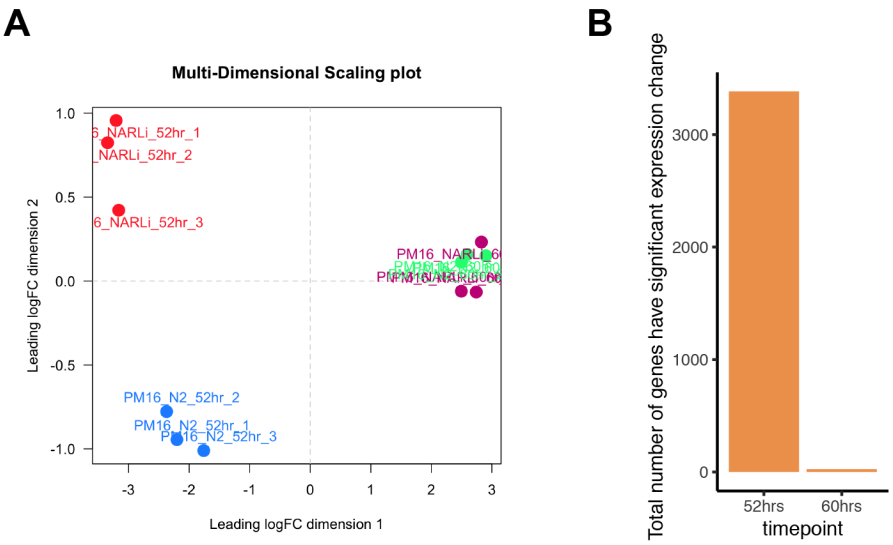

Figure 3.6. Transcriptome analysis of NARL*intron* and N2. A) Principal Component Analysis of N2 and NARL*intron* at 52 hours and 60 hours time point. These two timepoints were intentionally chosen to capture differences during spermatogenesis and oogenesis. RNA samples at 52 hours and 60 hours are separable and N2, NARL*intron* are separable only at 52-hour time point. B) N2 and NARL*intron* transcriptome are most different at 52-hour time point. the y-axis is the total number of differentially expressed genes between N2 and NARL*intron* at different time points.

Transcriptome analysis was also applied to NARL*intron* and N2 at two stages: 52 hours and 60 hours post eggs being laid. These two stages were selected to test the differential expression during spermatogenesis and oogenesis. N2 and NARL*intron* animals reach the L4 stage at 48-hour post egg, that is when spermatogenesis initiates. At 52 hours, spermatogenesis should be very active. At 60 hours, animals have entered the young adult stage and started oogenesis and making eggs. RNAseq analysis found that NARL*intron* and N2 have 3384 genes that are differentially expressed at 52-hour time point, which is significantly more than 25 differentially expressed genes at 60 hours (**Figure 3.6B**). This suggests that *nurf-1* intron SNP can affect a lot of genes during spermatogenesis. A further step can be taken to carefully look at the genes that are differentially expressed.

3.3 Methods

3.3.1 Strains

N2, CX12311 (N2*): *kyIR1(V, CB4856>N2), qgIR1(X, CB4856>N2)*, PTM66(NIL_{nurf-1}): *kyIR87(II, LSJ2>N2); kyIR1(V, CB4856>N2), qgIR1(X, CB4856>N2)*, PTM88 (ARL_{del}): *kyIR1 (V, CB4856>N2); qgIR1 (X, CB4856>N2); nurf-1(kah3) II; spe-9(kah132) I*, PTM416 ($NARL_{intron}$): *nurf-1(kah127) II, PTM229: dpy-10(kah82)II*

3.3.2 Sample preparation for RNAseq

RNAseq samples contains *nurf-1* 60bp deletion

CX12311, PTM66, PTM88, LSJ2 L4 hermaphrodites were picked to fresh NGM agar plates. Their adult progeny were bleached using alkaline-bleach solution to isolate eggs for synchronization. The eggs were washed with M9 buffer for three times and placed on a tube roller overnight. About 400 hatched L1 animals were placed on NGM agar plates and incubated at 20°C until they reach young adulthood. Young adulthood was defined as the time point when eggs were observed on assay plates. These worms were then harvested and washed for 3 times with M9 buffer. Samples were then kept frozen in -80 °C freezer until RNA extraction. Each strain and condition had 1 independent replicate on 3 different days.

RNAseq samples contain *nurf-1* intron SNP

N2 and PTM416 worms were synchronized through 3-hour egg laying. Worms were observed every hour after 46 hours until most of them reach the L4 stage, this is at timepoint 48 hours. Four hours later, spermatogenesis active worms were collected and kept frozen

in -80 °C freezer until RNA extraction. 12 hours later, young adult animals were collected and kept frozen in -80 °C freezer until RNA extraction. Each strain and condition had 3 independent replicates on the same day.

3.3.3 Transcriptome analysis

RNAseq and transcriptome analysis were performed as previously described[27]. Worms collected for RNAseq were proceeded to standard RNA isolation using Trizol. The RNA libraries for next-generation sequencing were prepared using an Illumina TruSeq Stranded mRNA kit (20020595) following its standard protocol. These libraries were sequenced using an Illumina NextSeq 500 platform. Reads were aligned using HISAT2 using default parameters for pair-end sequencing. Transcript abundance was calculated using HTseq or featureCount and then used as inputs for the SARTools (Varet et al., 2016). Within this R package, edgeR is used for normalization and differential analysis. N2* or N2 is treated as wild type. The genes showing significantly different expression ($\log_2$(fold) >1 or $\log_2$(fold) < −1, FDR adjusted p-value<0.01) were selected to perform Hierarchical Cluster analysis, and Principal Component Analysis.

3.3.4 Reciprocal heterozygosity test

Before test, I used CRISPR/Cas9 genome editing method created two strains carrying an early stop codon in the 2nd exon of *nurf-1* in NIL_{nurf-1} and ARL_{del}, naming NIL_{stop2} and ARL_{stop2}. These two strains were not able to reproduce at a normal rate, so I balanced them using a fluorescent marker. To create strains for reciprocal heterozygosity test, I crossed NIL_{stop2} to ARL_{del} and ARL_{stop2} to NIL_{nurf-1}. Their F1 L4 progenies were then singled to the assay plate for measuring egg laying rate as described in 2.3.4 except for different time

points were chosen: 0-24hours, 24-36hours, 36-48hours, 48-60hours, 60-72hours, 72-84hours. After assay, all singled worms were genotyped for validation purpose. Statistical test was performed at each time point using an unpaired two-tailed t-test between the following groups: NIL_{nurf-1} and *nurf-1(lof)/nurf-1(LSJ2 intron)*, NIL_{nurf-1} and *nurf-1(lof)/nurf-1(N2 intron)*, ARL_{del} and *nurf-1(lof)/nurf-1(LSJ2 intron)*, ARL_{del} and *nurf-1(lof)/nurf-1(N2 intron)*.

3.3.5 CRISPR $NARL_{intron}$ strain construction

Strain construction has three steps:

1. 2nd intron *dpy-10* replacement through co-conversion CRISPR/Cas9 editing:

   We designed four sgRNAs to target the

   5'- TCGATAATTATCCGTTTGT(GGG) -3' ,

   5'- TTGCATCATATCCCACAAA(CGG) - 3' ,

   5'- ACGGTAGCTCATGAAGAGA(AGG) -3'

   and 5'- TTCCGACGAATATAAGAAA(CGG) -3'

   protospacer/PAM sites to cut the 2nd intron and replace it with repair oligonucleotide: 5'- GTCTGTTAGAGATGCTATTAATGTCGATAATTATCgct accataggcaccacgagcgagATTCGTCGGAATTTAAGAAACTTGTGAATAATGT T -3' that contains a high efficient *dpy-10* sgRNA target sequence with NGAG PAM site that a Cas9 variant spCas9 can target: 5'- gctaccataggcaccacgag(cgag) -3'. I injected 50 ng/ul Peft-3::Cas9, 25 ng/ul dpy-10 sgRNA, 500 nM dpy-10(cn64) repair oligo, 10 ng/ul of each *nurf-1* intron sgRNAs, and 500 nM *nurf-1* intron repair oligonucleotide into N2 animals. Their decedents with a roller phenotype were then genotyped using the primers:

5’- GCAGGCCGGCCTTCGCGCCTGGGTAATACC -3’ and

5’- CGGCAGTTTTCGTCGTTCTG -3’

Followed by digestion with BanI (a restriction site created when deleting the second intron). I got a heterozygote worm that contains the replaced intron, but cannot homozygote this animal since its reproduction was impaired by the genome editing of *nurf-1*. To be able to proceed to the next editing, I genotyped and found a non-rolling animal that is heterozygotes due to recombination event. I then singled the progenies of the heterozygote animals and crossed it with a *nurf-1* GFP fluorescent marker strain and screen for balanced worms that contain the target intron replacement, this strain is named PTM366 *nurf-1(kah125)/oxTi924 II; kyIR1 (V, CB4856>N2); qgIR1 (X, CB4856>N2) X.*

2. 2nd intron LSJ2 allele replacement through co-conversion CRISPR/Cas9 editing:

   PTM366 cross NIL_{nurf-1} creates a heterozygote strain with one chromosome containing the intron *dpy-10* replacement and another chromosome containing the LSJ2 version of the intron SNP, which can be served as a repair template. I injected 50 ng/ul spCas9, 10 ng/ul dpy-10 sgRNA, 500 nM dpy-10(cn64) repair oligo. Unfortunately, I was not able to get the target repair through multiple rounds of screening. It may be that the dpy-10 sgRNA’s efficiency only applies to Cas9 but not for spCas9. So we designed two other sgRNA target sites 5’- ATctcgctcgtggtgccta(tgg) -3’ and 5’- TTCCGACGAATctcgctcg(tgg) -3’. I then injected 50 ng/ul Cas9, 10 ng/ul dpy-10 sgRNA, 500 nM dpy-10(cn64) repair oligo and 25 ng/ul for each *nurf-1* intron sgRNA. Progenies were then PCR genotyped to screen for LSJ2 allele at intron variation site using primers:

5’- GCAGGCCGGCCTTCGCGCCTGGGTAATACC -3’ and

5’- CGGCAGTTTTCGTCGTTCTG -3’

and heterozygote at 60bp deletion site using primers:

5’- CGACAAAAAGTTGATAGACG -3’

5’- CATCTTCATAATTCCAACGGAAACCAAG -3’

After screening, the target genotype can be homozygote, this strain was then named as PTM410 ARL$_{intron}$ k*yIR1 (V, CB4856>N2); qgIR1 (X, CB4856>N2); nurf-1(kah127)II.*

3. ARL$_{intron}$ backcross to N2 background

ARLintron was then backcrossed to N2 background using an RFP fluorescent *nurf-1* marker strain for 4 generations and the *npr-1* and *glb-5* sites were genotyped as described previously[21] to make sure that they carry the N2 alleles.

CHAPTER 4. Two Short Major *nurf-1* Isoforms Work Antagonistically During Gametogenesis

4.1 Introduction

NURF-1 is the *C. elegans* ortholog of human Bromodomain PHD-finger Transcription Factor (BPTF) and *Drosophila* NURF301, the largest subunit of Nucleosome Remodeling Factor (NURF) complex, which is an essential component in chromatin remodeling[28]. BPTF was first discovered as Alz-50 Clone1 (FAC1) protein[29]. In the human fetal brain, FAC1 is an 810 amino acids long protein that is reactive to an antibody designed to recognize Alzheimer's disease neurofibrillary pathology[29]. But BPTF actually encodes a protein product that contains 2781 amino acids[30]. As a subunit of NURF complex, it cooperates ATPase ISWI to slide nucleosomes in 10bp bidirectional step movement until reach thermodynamically stable state, or the end of the barriers constructed by DNA binding factors or other nucleosomes (**Figure 4.1**) [31], [32].

Researches on BPTF often focus on the long isoform, which includes all functioning domains: ISWI interacting domain DDT; histone marks recognition domains PHD finger and bromodomain; other transcription factor binding or DNA binding domains (**Figure 4.1)**. But human BPTF has 25 computationally-predicted isoforms as of today. 21 of these isoforms cover the full length of the gene and are derived from cDNA and EST data. Only recently, 4 other short N-terminal and C-terminal isoforms are predicted on NCBI through eukaryotic genome annotation pipeline, each with a subset of the functional domains

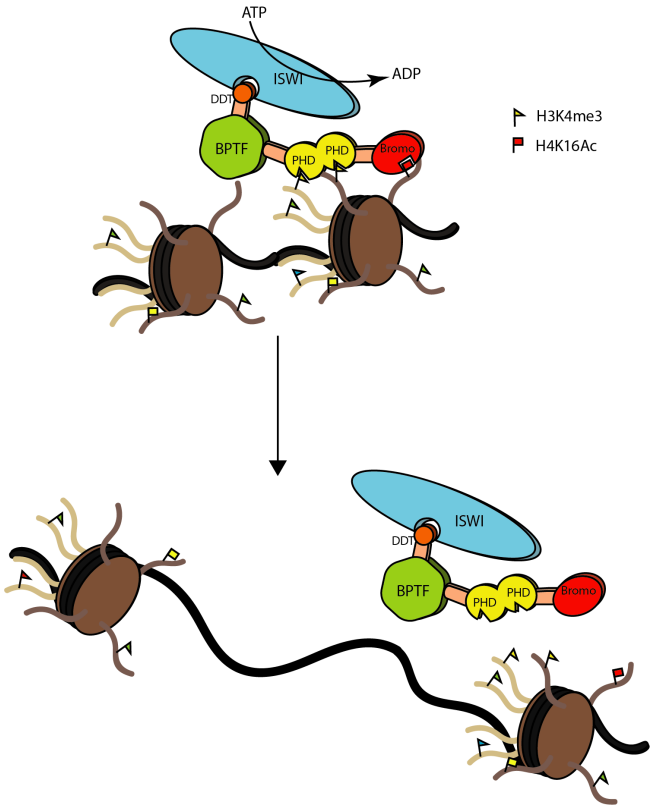

Figure 4.1. BPTF working model. BPTF contains multiple functional domains: DDT domain is essential for interaction with ATPase ISWI, two C-terminal PHD domains are important for H3K4me3 recognition and the bromodomain can recognize H4K16Ac. The reported mechanism of NURF function is through histone marks recognition and utilize ATP to slide out the compact DNA thus to make those regions accessible for transcriptional factors.

(**Figure 4.2**)[30]. *In vitro* studies suggest that the BPTF interacts with various transcription factors, adaptor proteins etc. to regulate diverse biological functions like tissue development, stress response, reproduction and cell proliferation [28]. For example, BPTF interacts with Myc-Associated Zinc finger (ZF87/MAZ) transcription factor to regulate neurodegeneration and interacts with human Kelch-like ECH-associated protein 1 (hKeap1) to regulate stress response [33], [34]. In addition to its regulatory role, NURF also affects certain disease progression progresses. It's deletion

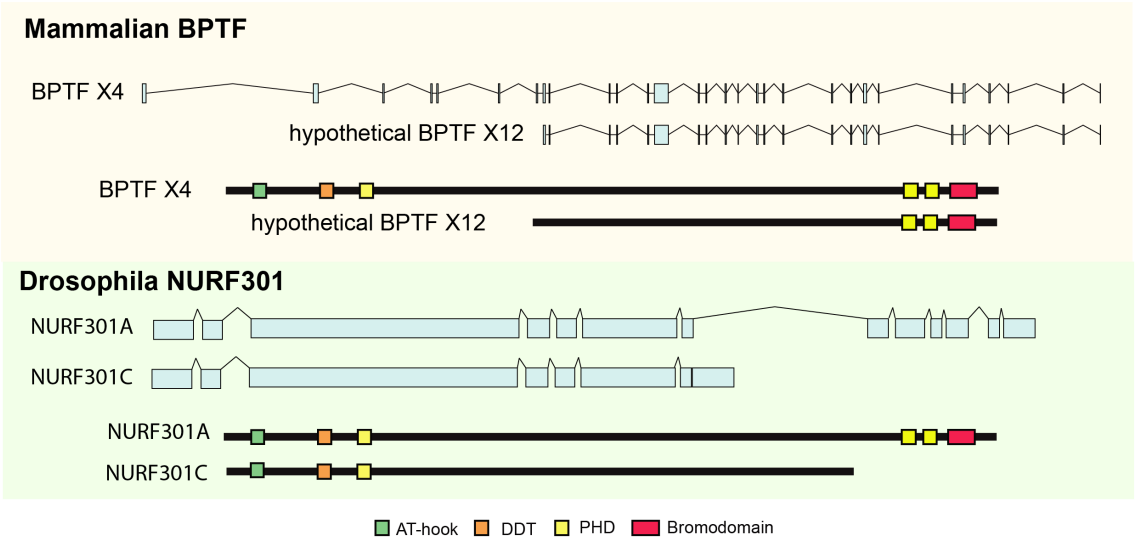

Figure 4.2 Predicted isoforms of *nurf-1* orthologs. Most mammalian BPTF studies focus on the long form, only recently a short C-terminal form that does not contain ISWI-interaction domain DDT was predicted. In *Drosophila*, two forms were verified to exist: a long form and an N-terminal form that does not have histone marks recognition domains.

mutation, frame-shift mutation, promoter duplication, copy number variant, and gene fusion were reported in various diseases like neuroblastoma, lacrimal gland cancer, colorectal carcinoma, leukemia etc. [35]–[41].

Study on NURF301, the Drosophila ortholog of *nurf-1*, and its mutations revealed its importance in transcription activation, homeotic gene expression, stress response, male X chromosome morphology, pupation and reproduction [42]–[44]. In addition to the long form NURF301, a short isoform NURF301C that only contains the N-terminal part of the gene was also studied in detail (**Figure 4.2**). Kown et. Al reported that NURF301C was required for gametogenesis and only 20% of NURF301's target will be affected if this short isoform was removed[44], suggesting 80% of the gene regulatory functions was performed by other transcripts. I have two takeaways from this result. First, this NURF301C transcript

does not contain PHD, Bromodomains that are essential for histone recognition. The regulatory role of NURF301C suggests there exist other nucleosome remodeling machinery for NURF301. Second, NURF301C only regulates part of the genes NURF301 affects, suggesting diverse function of NURF301 isoforms.

In *C. elegans*, the *nurf-1* gene is predicted to encode 16 distinct isoforms that can be classified into 9 representative forms with start or stop site differences (**Figure 4.3A**)[45]. Within each group, there are minor (6-9bp) alternative splice sites differences. The 9 representative isoforms are either long forms, short N-terminal forms or short C-terminal forms, each contains all or subset of the functional domains (**Figure 4.3B**). In 2006, Erik Andersen et. al. first characterized 5 *nurf-1* isoforms through cDNA clones, RT-PCR and 5'RACE[46]. They also isolated two classical alleles *nurf-1(n4293)* and *nurf-1(n4295)* from EMS-mutagenized animals. *nurf-1(n4293)* removes the N-terminal part of the gene that affects *nurf-1.a* and *nurf-1.b* but not the other *nurf-1* variants. *nurf-1(n4295)* removes part of every *nurf-1* variants except for *nurf-1.b*. *nurf-1(n4293)* worms were sterile but able to suppress synMuv (synthetic multivulva) phenotype of lin-15AB mutant animals. *nurf-1(n4295)* animals were grossly wildtype but could not suppress lin-15AB mutation induced synMuv [46]. These findings suggest the diverse function of *nurf-1* isoforms.

I hypothesize that the reason why *nurf-1* was repeatedly targeted during evolutionary is that different isoforms have different effects on fitness proximal traits.

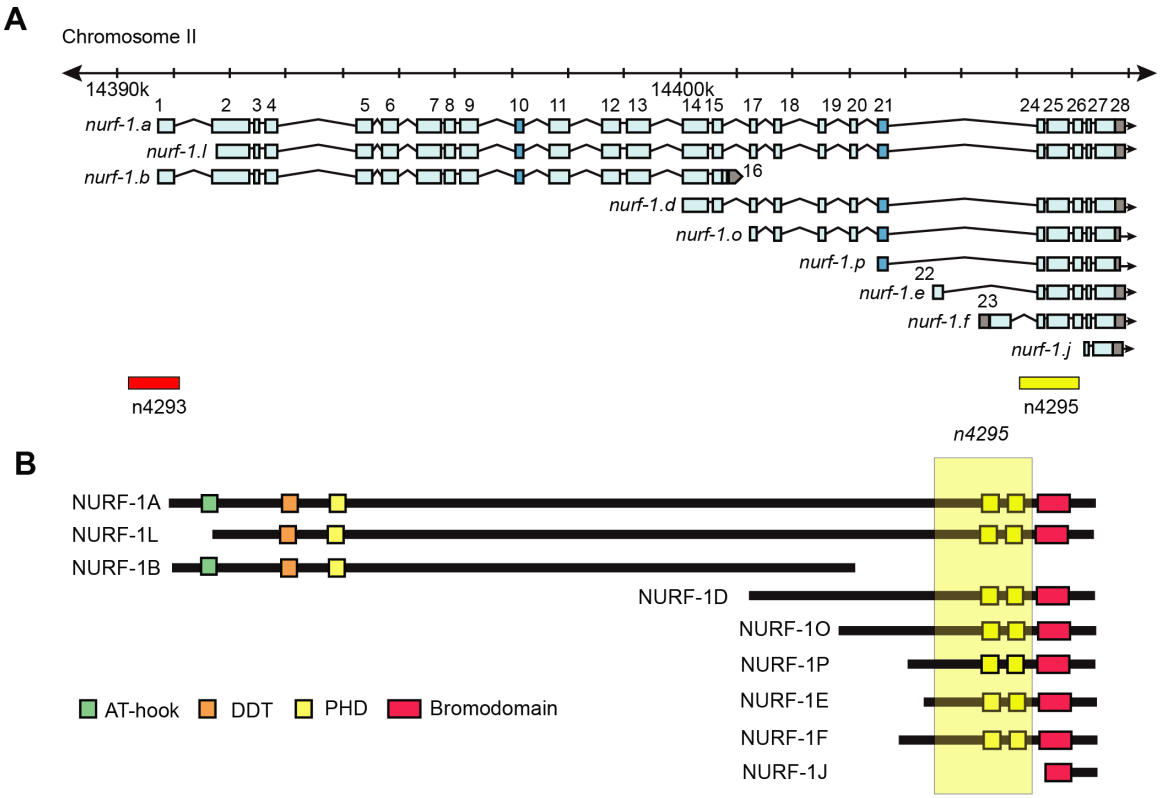

Figure 4.3. *nurf-1* encodes 16 isoforms that contain different functional domains. A) 8 representative forms of 16 *nurf-1* isoforms. Exon 10 and exon 21 are labeled in dark blue since there exist alternative splicing events. There are 8 predicted start and two stop sites for *nurf-1*. **B)** Protein domains for the 9 representative isoforms. NURF-1A is the long form that contains all functional domains, NURF-1B does not have the C-terminal PHD domains and Bromodomain for histone marks recognition which NURF-1D, O, P, E, and F do not have the DDT domain that's essential for interacting with ISWI. NURF-1L is a special form that was recently predicted that lacks the AT-hook that can bind DNA. NURF-1J only contains part of the bromodomain.

Furthermore, I hypothesize the 60bp deletion and the intron SNP affect different *nurf-1* isoforms that caused the phenotypic divergence of N2 and LSJ2.

4.2 Results and Discussion

4.2.1 Promoter-driven GFP expression assay rules out two isoforms of nurf-1

Before test the function of different isoforms, I decided to verify the existence of those predicts isoforms by using a 500bp promoter region before each predicted start site to drive the expression of GFP protein. This method is based on *C. elegans*'s tolerance to extra-chromosomal arrays. By injecting a plasmid carrying a functional promoter, GFP coding sequence and a universal 5'-UTR to *C. elegans* germline, its next generation will be able to express GFP at cells where this specific promoter is active. I found promoters before the 1st, 2nd, 14th, 17th, 22nd, 23rd exon were able to drive the expression of GFP

Table 4.1. GFP expression drove from different *nurf-1* promoters.

the promoter of exon #	observed GFP expression	observed neuronal expression	observed pharynx expression	observed intestine expression
1	yes	yes	yes	yes
2	yes	yes	yes	yes
14	yes	yes	yes	yes
17	yes	yes	yes	yes
21	no	n/a	n/a	n/a
22	yes	no	yes	no
23	yes	yes	yes	yes
27	no	n/a	n/a	n/a

(**Table 4.1**), leaving out promoter 21 and promoter 27 that drive expression of *nurf-1.p* and *nurf-1.j* respectively. These two isoforms were then excluded from later discussion. When observing the promoter-driven expression of GFP, I noticed that each promoter can result

Figure 4.4. Distinct expression pattern for different *nurf-1* promoters. Representative figures showing the expression pattern of different *nurf-1* promoters. Exon 1 promoter can drive neuronal and gut expression etc. and exon 2 promoters can drive expression of a larger set of neurons and gut, while exon 22 promoter can only drive the expression in the pharynx.

in very diverse expression as shown in **Figure 4.4**, suggesting cell-specific expression of isoforms. This further supports my hypothesis that *nurf-1* isoforms have diverse functions.

4.2.2 Nanopore sequencing reads support four major nurf-1 forms and a new nurf-1.q isoform

Recently, Nanopore-based sequencing technology has been used to generate long (>5kb) reads of mRNA molecules or cDNA molecules generated from mRNA, including RNA isolated from a mixed stage culture of *C. elegans*. These long reads can be used to resolve the major transcripts of *nurf-1*. From the recently published *C. elegans* dataset, we identified long reads that aligned to *nurf-1* and classified each as matching one of the five major transcripts (**Figure 4.4**). 3 reads matched the *nurf-1. a* transcript, 25 reads matched the *nurf-1.b* transcript, 7 reads matched the *nurf-1.d* transcript, 6 reads matched the *nurf-1.f* transcript and 50 reads matched a new *nurf-1.q* isoform. The *nurf-1.q* transcript initiates

from the 14th exon like *the nurf-1.d* isoform, however, it also undergoes the alternative splice event of the *nurf-1.b* isoform, terminating after the 16th exon. This transcript has not been described before, however, is well-supported by the Nanopore reads. We also analyzed RNAseq data we have generated using Illumina short-read sequencing to compare with the Nanopore analysis. Both technologies largely agreed with each other. Clipped reads containing spliced leader 1 sequence (trans-spliced to transcriptional start sites) [47] were only found at the 1st and 14th exon. A large number of reads also aligned to the 23rd exon, consistent with these three sites serving as the primary locations of transcript initiation. The coverage of the 14th, 15th, and 16th exons was also much higher than the 1st – 13th and 13th – 28th exons, consistent with a *nurf-1.q* transcript. We could also use this data to identify small variations in splicing in the 10th, 16th, and 23rd exons, resulting in the loss/inclusion of 2 amino acids, 2 amino acids, and 3 amino acids respectively (shown in blue in **Figure 4.5A**). For simplicity, we ignore the subtle variations in transcripts these alternative splicing events would create, subsuming this complexity into the five major isoforms. This variation accounts for part of the additional complexity of the WormBase predictions. For other WormBase-predicted transcripts (e.g. *nurf-1.e, nurf-1.g, nurf-1.h, nurf.1.l, nurf-1.m, nurf-1.o,* and *nurf-1.p*), we will ignore in the context of this paper for simplicity as well as less support for these transcripts from the Nanopore or Illumina sequencing. The predicted protein products produced by the five major transcripts is shown in **Figure 4.4C**.

4.2.3 Western blot provides protein evidence of two nurf-1 *isoforms*

While each of the five major transcripts is transcribed, that does not necessarily mean they are translated into stable protein products. Antibodies against BPTF have been raised

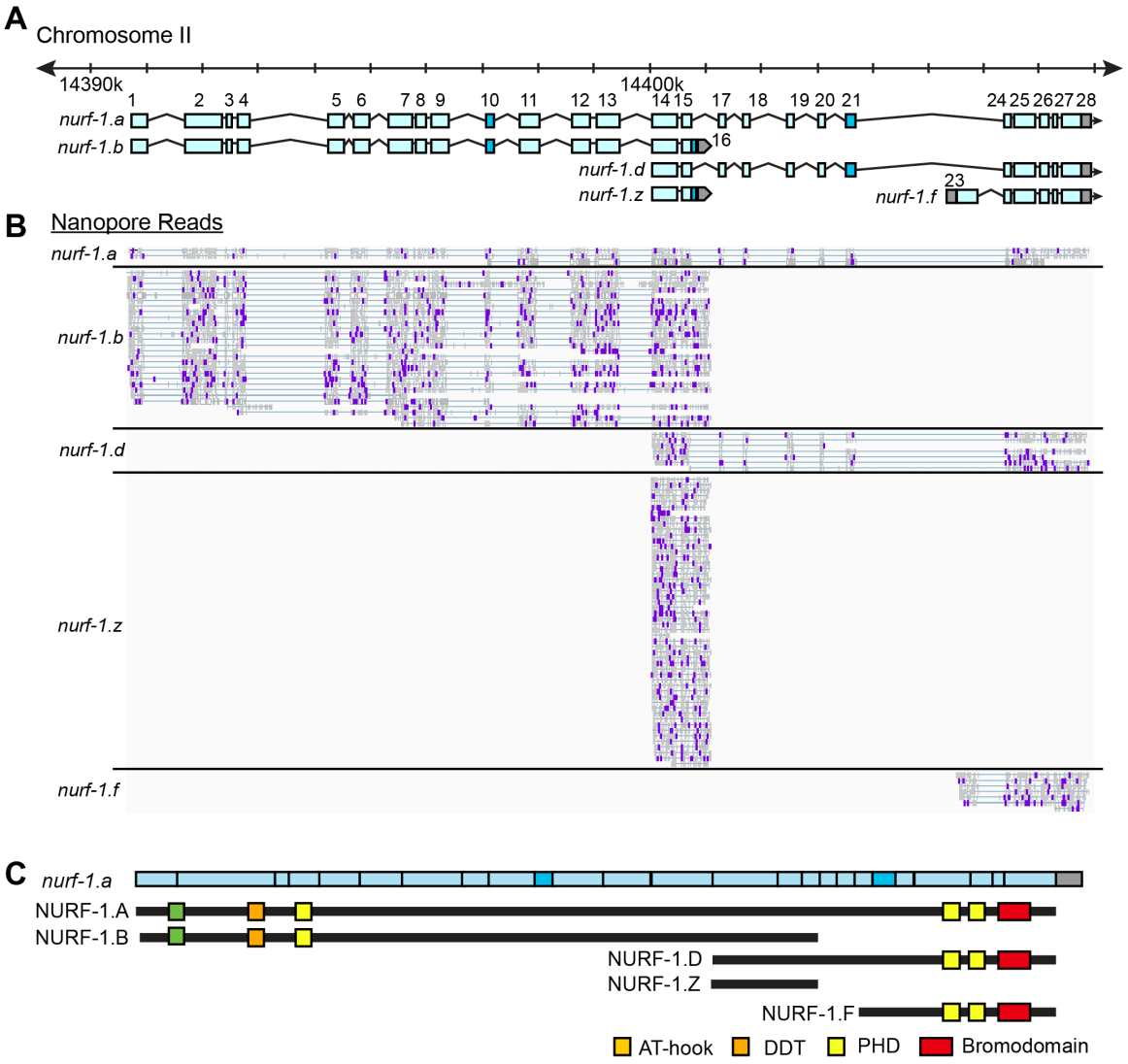

Figure 4.5. Nanopore reads validation of *nurf-1* isoforms. A) nurf-1 isoforms supported by nanopore reads. **B)** Original nanopore reads mapped to nurf-1 gene region. **C)** Nanopore supported nurf-1 isoforms' protein and functional domains prediction.

against recombinant protein from humans and mouse, however, these could not recognize proteins of the appropriate size in *C. elegans* protein lysate. To facilitate analysis of NURF-1 proteins, I used CRISPR/Cas9 to edit the endogenous *nurf-1* locus to fuse two distinct epitope tags, an HA tag, and a FLAG tag, just prior to the stop codons in the 16th and 28th exon (**Figure 4.5A**). This epitope-tagged strain PTM420 grew slower and laid fewer eggs

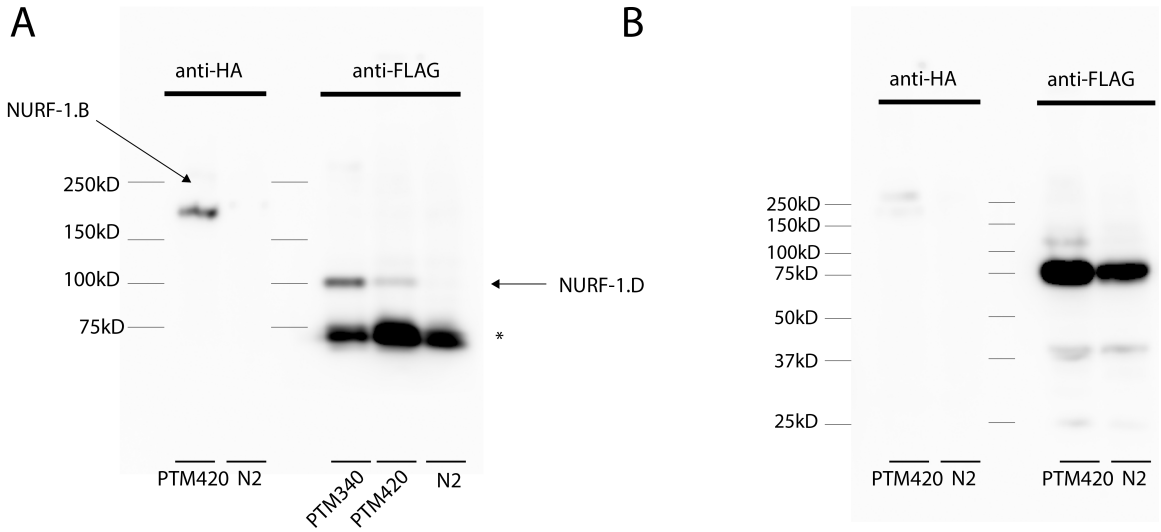

Figure 4.6. Western blot against *nurf-1* isoforms with epitope tags. A) Western blot for 6% gel samples with anti-HA and anti-FLAG antibody. 6% gel was targeted to resolute proteins have molecular weight more than 75kD, that is isoform *NURF-1.A, NURF-1.B,* and *NURF-1.D.* The left panel showed detected band of *NURF-1.B* and the right panel showed the detected panel of *NURF-1.D.* No evidence of *NURF-1.A* was observed. **B)** Western blot for 10% gel samples with anti-HA and anti-FLAG antibody. 10% gel was targeted to resolute proteins have molecular weight lower than 75kD.

than N2 control, suggesting that these tags partially inhibited the function of the protein. Nevertheless, we decided to use this strain to determine if proteins were produced for the five predicted isoforms. *NURF-1.A, NURF-1.B, NURF-1.D, NURF-1.F,NURF-1.Q* encode proteins of size 252.6kD, 185.76kD, 92.3kD, 57.89kD, 25.81kD, respectively. Western blot using anti-HA antibody will be able to resolute if *NURF-1.B* and *NURF-1.Q* exist, and anti-FLAG antibody can detect the protein bands for *NURF-1.A, NURF-1.D, and NURF-1.F* if exist. Western blot result is shown in **Figure 4.6** suggest the existence of *NURF-1.B* and *NURF-1.D.* From Nanopore sequencing, I expect *NURF-1.A* harder to resolute due to the low number of aligned reads. So it is possible that *NURF-1. A* protein product exists but my western blot does not have the power to detect it. Similarly, I also expect to see a

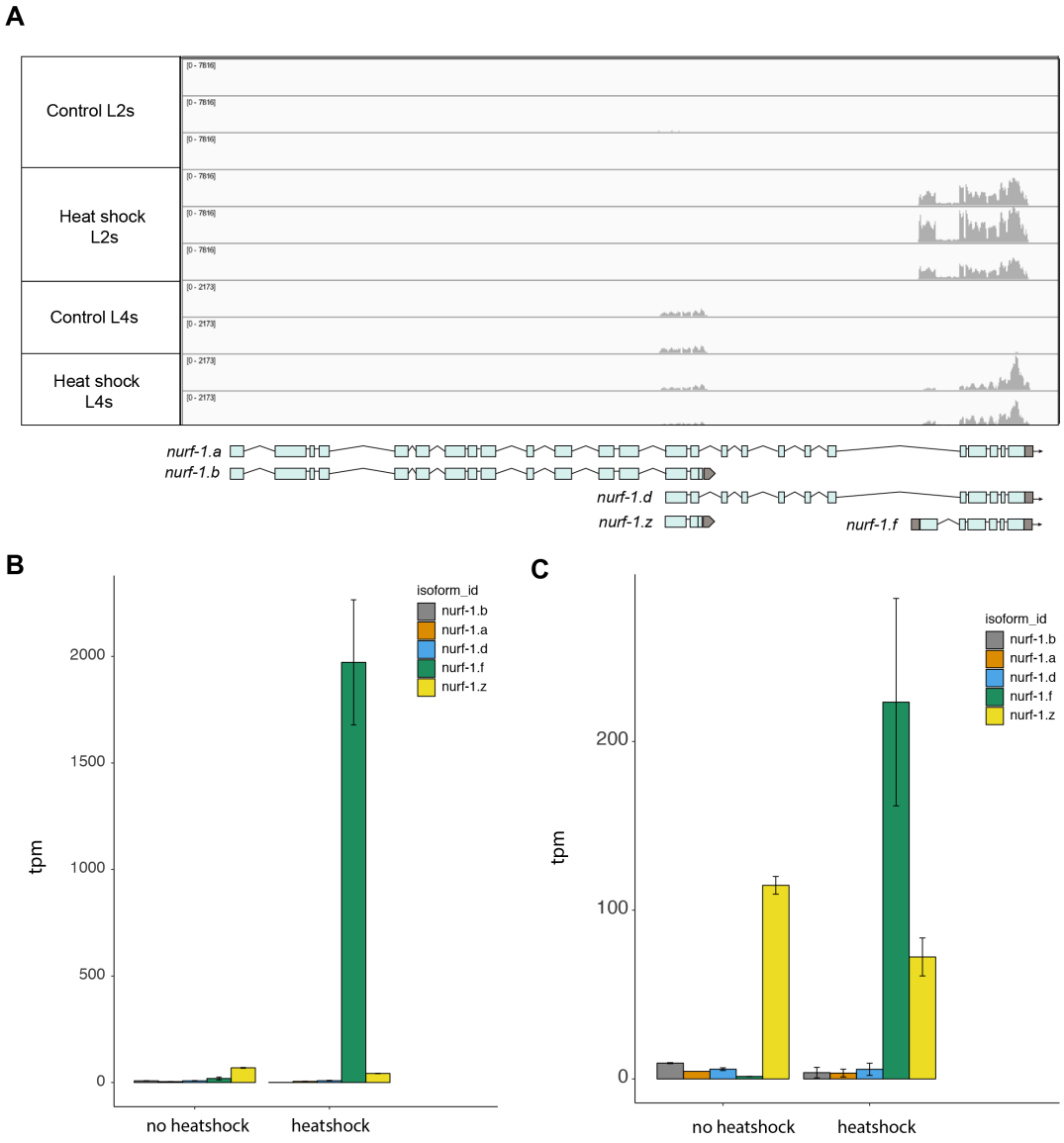

Figure 4.7. *nurf-1* transcripts differential expression after heat shock. A) IGV view of bam file of RNAseq data from control and heat shocked worms. **B)** RNAseq data from Guisbert et. al[48] suggest significantly elevated expression of *NURF-1.F* isoform in response to heat shock. y-axis suggest transcripts per million reads. **C)** RNAseq data from Brunquell et. al also suggest significantly elevated expression of *NURF-1.F* isoform in response to heat shock. y-axis suggest transcripts per million reads.

significantly brighter band of *NURF-1.Q* if this isoform exists, but turn out to see none. I conclude that *nurf-1.q* RNA does not translate to protein products.

4.2.4 Heat shock specifically upregulates the F isoform of nurf-1

No evidence of *NURF-1.F* isoform was found in my western blot. This may due to the strong non-specific band at ~75kD as *NURF-1.F* size is 57.89kD. I changed to another anti-FLAG antibody to avoid non-specific bands but still was not able to resolute *NURF-1.F* isoform. Guisbert et. al. reported overexpression of *NURF-1* under heat shock (Figure 4.6A) [48]. I analyzed their RNAseq data using kallisto[49], a pseudo-alignment program for quantifying the relative abundance of different transcripts and found *NURF-1.F* transcript was significantly upregulated after heat shock (**Figure 4.7B**). The same trend was also found in another dataset that performed heat shock on worms at the L4 stage (**Figure 4.7C**) [50].

In order to examine why *NURF-1.F* expression increased after heat shock, I created a strain *kah144* that contains a loss of function (lof) deletion that only affects *nurf-1.f* (**Figure 4.8A**) in the epitope tag background. To test the heat shock response of *NURF-1.F* expression, I performed western blot for *kah144* and used N2 and PTM420 as negative controls. Results verified elevated *NURF-1.F* protein product after heat shock (**Figure 4.8A** and **B**).

4.2.5 Loss of nurf-1.f *isoform does not affect fitness proximal traits*

To test the fitness proximal traits for *nurf-1.f (lof)*, I created strain contains a STOP codon in *nurf-1.f* in N2 background STOP Isogenic Lines 23 (SIL_{23}) (**Figure 4.9A**). After measuring its fecundity and growth, I found no significant difference for *kah11* and N2 (**Figure 4.9B and C**). This is not surprising since the expression of *NURF-1.F* could only be detected after heat shock. I propose *nurf-1.f* may involve in heat shock response pathways.

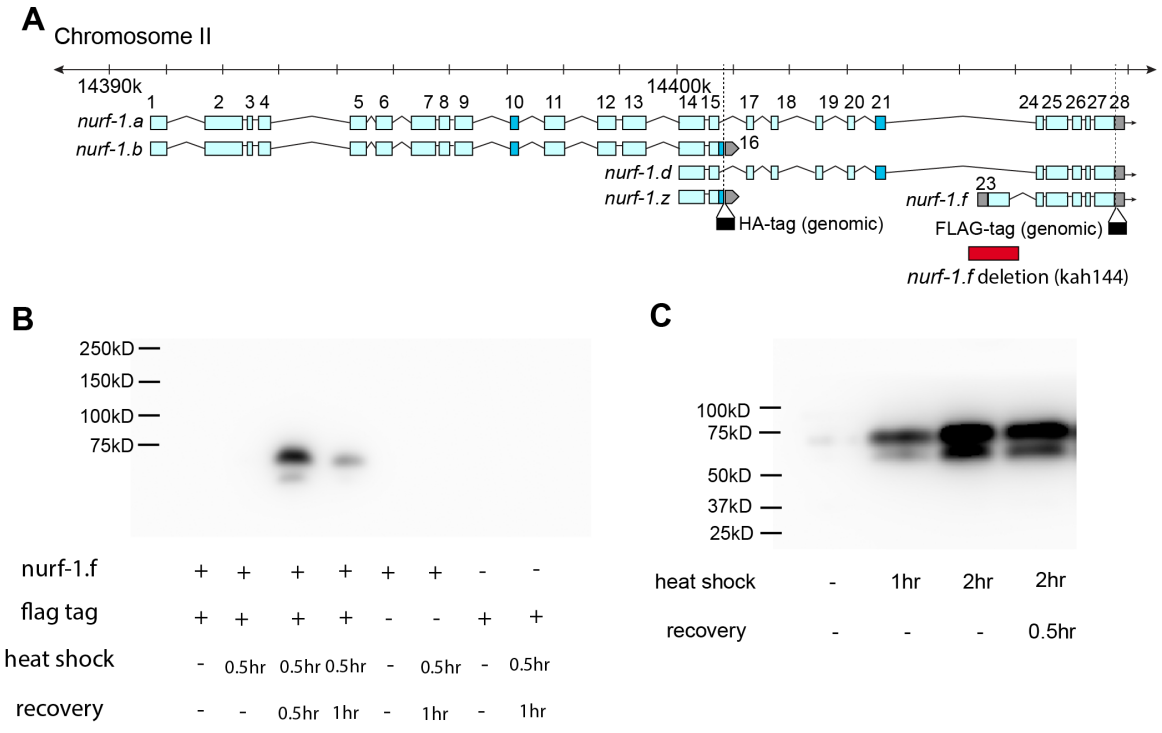

Figure 4.8. *NURF-1.F* isoform protein expression is up-regulated after heat shock.

A) *kah144* strain has a deletion of 23rd exon shown in red. **B)** Heat shock response analysis of N2, PTM420, and *kah144*. Negative controls N2 and *kah144* didn't show any clues of *NURF-1.F*. *NURF-1.F* is only detectable after 30 minutes heat shock with recovery. **C)** Heat shock response does not require recovery. Previous heat shock western blot only allowed 30 minutes for F isoform to express. Here I show the longer the heat shock, the more F isoform express.

4.2.6 nurf-1.f *transcriptome regulation in response to heat shock is relatively subtle*

To determine if this isoform had an effect on transcription in response to heat shock, I performed RNAseq on N2 and f deletion animals at 0 hours, 2, and 4 hours after heat shock as well as 2 hours of heat shock followed by 2 hours of recovery. Overall, most of the heat shock response was the same in the two strains, indicating that the NURF1.F is not a major regulator of heat shock response. However, we did observe a significant change of 4 hours

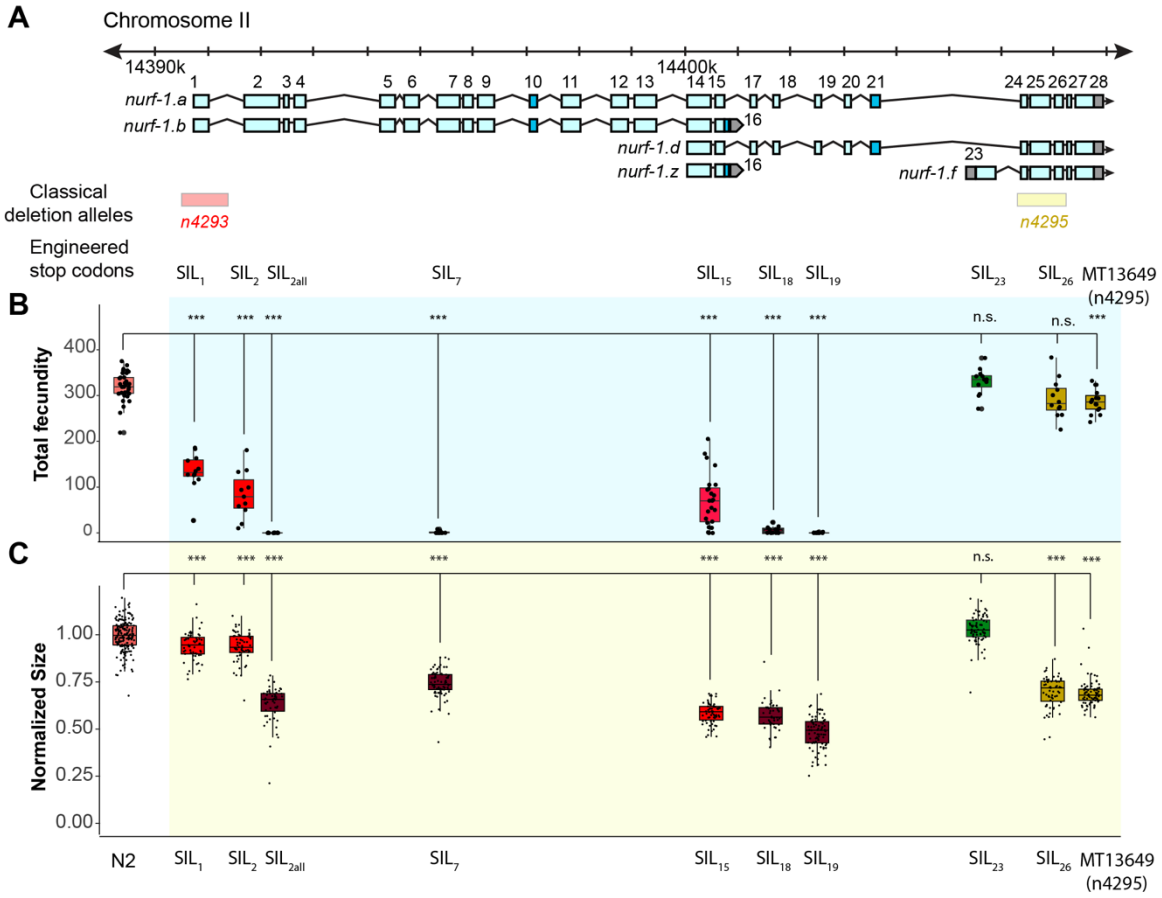

Figure 4.9. STOP Isogenic Lines (SILs) reproduction and growth.A) Positions of classical deletion alleles *n4293, n4295* and STOP codon lines. **B)** Total fecundity of all SILs and *n4295. the* y-axis is the total number of eggs laid each animal. *** indicates a significant difference at p-value<0.01 level. n.s. indicates no significance at p-value<0.1 level. C) Growth analysis of SILs and *n4295.*Animals' size was normalized to the average size of N2. *** indicates a significant difference at p-value<0.01 level. n.s. indicates no significance at p-value<0.1 level.

after heat shock. Few genes were different at 0 or 2 hours but by 4 hours of heat shock or after two hours of recovery, approximately 200 genes showed a change in expression (**Figure 4.10B**). Analysis of transcriptional response indicated a strong correlation between genes affected after four hours of heat shock and genes affected after two hours of heat shock followed by two hours of recovery with R^2 score of 0.4421(**Figure 4.10C**). We

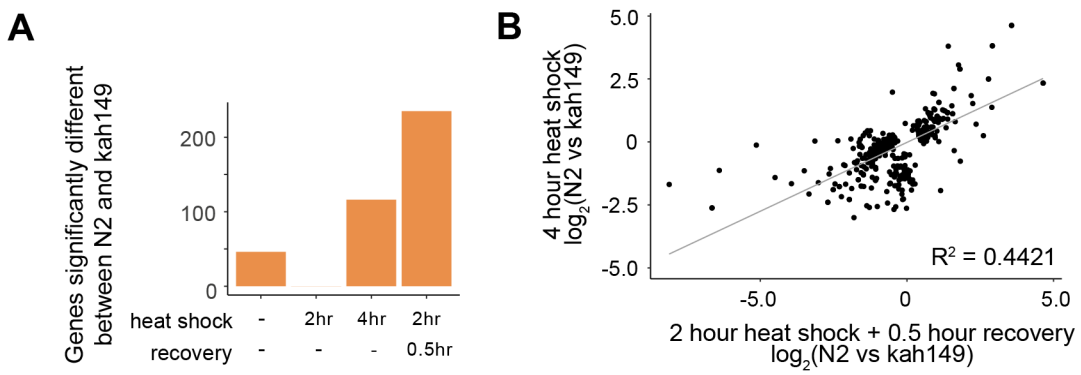

Figure 4.10. Transcriptome analysis of heat shock response.A) Total differentially expressed genes between N2 and *kah149* (23rd exon deletion). Animals were compared at no heat shock, heat shock 2 hours, heat shock 4 hours and 2hour heat shock with 30 minutes recovery. Heat shock with recovery is the condition that N2 and *kah149* are most different. **B)** For all differentially expressed genes in 4hour heat shock and 2hour heat shock plus 30 minutes recovery, their $\log_2$ fold change in these two conditions are positively correlated. ($R^2 = 0.4421$).

conclude that the F isoform is specifically up-regulated by heat shock and plays a modulatory role in determining the long-term transcriptional response to heat shock.

4.2.7 STOP Isogenic Lines (SILs) affecting the N-terminal isoforms phenocopied the classical allele n4293

To study the function of all other isoforms, I created stop isogenic lines (SILs) that carry CRISPR-induce STOP codon replacement for different exons: exon 1, exon 2, exon 2all, exon 7, exon 15, exon 18, exon 19 and exon 26. For simplification issue, I refer these strains as SIL_1, SIL_2, SIL_{2all}, SIL_7, SIL_{15}, SIL_{16}, SIL_{18}, SIL_{19} and SIL_{26} (**Figure 4.9A**). Two SILs were created for exon 2 to include analysis of recent Wormbase predicted *nurf-1.l* isoform (**Figure 4.2A**). The four stop codons in the 1st, 2nd, and 7th exons are predicted to result in loss-of-function mutations in the A and B isoforms and should phenocopy the *n4293*

allele. This prediction was mostly confirmed; each of the four mutations had a dramatic effect on reproduction and eventually led to sterility within 3-4 generations, although the severity varied between them (**Figure 4.9B**). The stop codon in the 1st exon (kah90) and the first stop codon in the 2nd exon (kah91) reduced total fecundity in the first generation (from ~300 to ~100) and resulted in a slight change in growth rate (**Figure 4.9C**). While homozygous animals were able to survive and lay a small number of eggs, their offspring were sick – many dying before or after hatching, and the few that survived to adulthood were often sterile or laid even fewer eggs until the line was extinguished, The 2nd stop codon in the 2nd exon, named as 2all and the stop codon in the 7th exon had more dramatic effect. These mutant animals laid almost no eggs and showed a larger decrease in growth rate (**Figure 4.9B** and **C**). What is the reason for the difference in severity in these mutants? One possibility is a difference in frequency of translational read-through of each stop codon, which is interpreted as sense codons at a low frequency[51]. Another possibility is that this difference has a biological basis due to the presence of a 2nd translational initiation site (i.e. one transcript produces two protein isoforms). The 2nd ATG codon in the A/B isoform occurs at the 122nd amino acid, downstream of both of the 1st two stop codon edits.

4.2.8 SIL_{26} phenocopied the classical allele n4295

We next analyzed the stop mutation in the 26th exon near the C-terminus, at a position within the *n4295* deletion, expecting this mutation to phenocopy the *n4295* deletion. Indeed, this mutant had a very similar phenotype, resulting in grossly wildtype animals with a minor difference in total fecundity and quantitative effect on reproductive timing and growth rates (**Figure 4.9B, C** and **4.11A**).

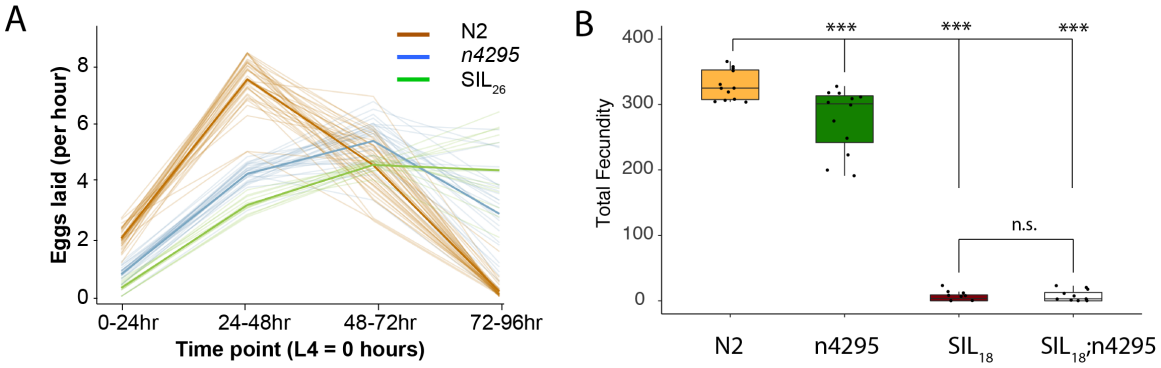

Figure 4.11. Egg laying rate analysis of SIL_{26} phenocopies *n4295* and n4295;SIL_{18} double mutant phenocopies SIL_{18}.A) Egg laying rate of N2, *n4295* and SIL_{26}. Both *n4295* and SIL_{26} affect the very C-terminal of *nurf-1* and show later reproductive timing phenotype. B) Total fecundity of N2, *n4295*, SIL_{18} and n4295;SIL_{18} double mutant. *n4295* had lower fecundity than N2, SIL_{18} and n4295;SIL_{18} double mutant had almost no fecundity. *** indicates a significant difference at p-value<0.01 level. n.s. indicates no significance at p-value<0.1 level.

4.2.9 nurf-1.d *is also essential for C. elegans*

We engineered stop codons in the 18th and 19th exon that predicted to create loss-of-function alleles the A and D isoforms. Unexpectedly, however, we found that both SIL_{18} and SIL_{19} homozygotes were almost completely sterile within the first generation (**Figure 4.9B**), suggesting that the *n4295* mutant is not a loss-of-function allele for the A and/or D isoforms. An alternative hypothesis is the A or D isoform could counterbalance negative effects of the F isoform so that loss of all three isoforms in the *n4295* mutant perturbs reproduction while the loss of just the A and D isoform in SIL_{18} and SIL_{19} results in sterility. However, we could exclude this possibility as the double mutant SIL_{18}; *n4295* was sterile like the *kah96* single mutant (**Figure 4.11B**).

Finally, a stop codon in the 15th exon resulted in an intermediate an intermediate phenotype. This stop codon is predicted to result in loss of function mutations in the A and D isoforms but probably not perturbs the B isoform, as it retains 1548 of 1621 amino acids.

4.2.10 Compound heterozygotes study suggest the long isoform nurf-1.a is not necessary for C. elegans

This genetic analysis is consistent with a number of possibilities: 1) the A isoform is essential, 2) the A and B isoform are both essential, 3) the B and D isoform are both essential, or 4) the A, B, and D isoform are all essential. In order to distinguish between these possibilities, I created compound heterozygotes between three of these stop mutation alleles (**Figure 4.12A**). I first verified that each of these mutations was recessive by measuring the fecundity of heterozygote animals (**Figure 4.12B**), demonstrating that a single copy of each isoform can rescue wildtype function. By combining independent mutations, we could specifically remove the long A isoform from the cells. For example, the SIL_{2all} mutation is predicted to ablate both the A and D isoforms, while the SIL_{18} mutant is predicted to ablate the A and B isoform. The SIL_{2all} /SIL_{18} compound heterozygotes are predicted to encode a single unaffected copy of the B isoform (from the *kah96* haplotype), a single unaffected copy of the D isoform (from the SIL_{2all} haplotype), but zero unaffected copies of the A isoform. If the A isoform was essential, we would expect the compound heterozygote animals to be sterile or with severe effects on fecundity. However, this animal was largely wild-type, suggesting that the full-length A isoform is not essential (**Figure 4.12B**). The SIL_{2all}/ SIL_{18} compound heterozygotes showed similar results. These animals are predicted to encode one unaffected copy of the D isoform, one truncated copy of the B isoform, and zero unaffected copies of the A isoform. These

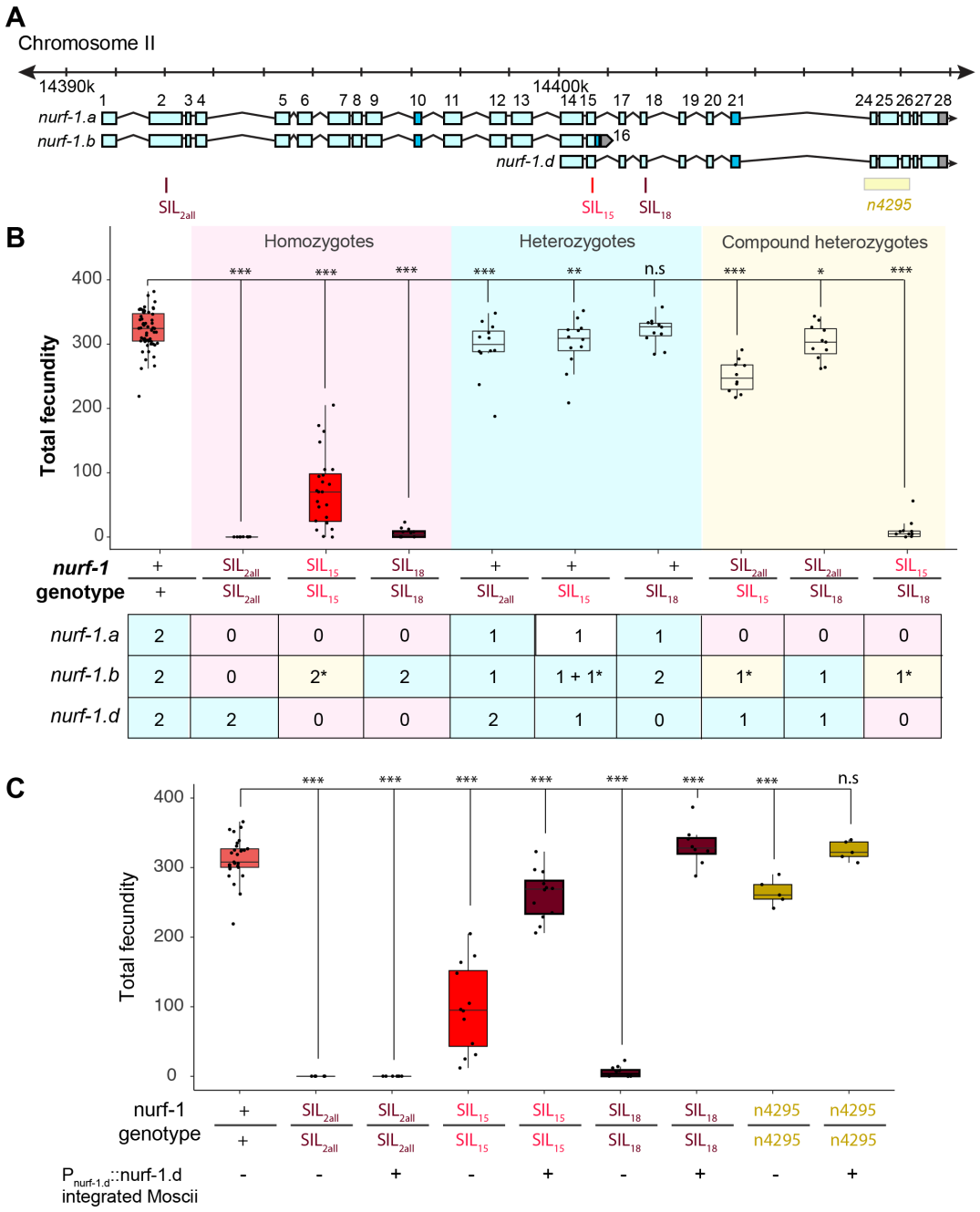

Figure 4.12. Compound heterozygotes and P*nurf-1.d::nurf-1* integrated rescue of SILs' fecundity. A) Relative positions of SIL_{2all}, SIL_{15}, SIL_{18} and *n4295* in *nurf-1*. B) Total fecundity for SIL_{2all}, SIL_{15}, SIL_{18} heterozygotes, and compound heterozygotes. *** indicates a significant difference at p-value<0.01 level. ** indicates a significant difference at p-value<0.05 level. * indicates significant difference at p-value<0.1 level. n.s. indicates no significance at p-value<0.1 level. C) Total fecundity of SIL_{2all}, SIL_{15}, SIL_{18}, and rescue strains have integrated P*nurf-1.d::nurf-1*. *** indicates a significant difference at p-value<0.01 level. * indicates significant difference at p-value<0.1 level. n.s. indicates no significance at p-value<0.1 level.

animals were largely wildtype, with a small reduction in total fecundity (**Figure 4.12B**). Our interpretation of this data is that the A isoform is not essential and the truncation of the B isoform slightly perturbs its function.

4.2.11 nurf-1.b *and* nurf-1.d *are both essential for C. elegans*

Finally, we analyzed SIL_{15} / SIL_{18} compound heterozygotes. These animals are predicted to encode zero unaffected copies of the D isoform, one unaffected copy of the B isoform, and zero unaffected copies of the A isoform. These animals were essentially sterile. Taken together, we interpret this data as indicating that the B and the D isoform are both essential for *C. elegans* reproduction.

4.2.12 Rescue assay further support nurf-1.d *is essential*

To confirm that the B isoform is essential, I also created a transgenic strain containing an integrated construct driving a *nurf-1.d* cDNA from its endogenous promoter. This transgene could fully rescue the fecundity phenotype of the SIL_{18} and *n4295* mutants, and partially rescue the fecundity phenotype of the SIL_{15} mutants (**Figure 4.12C**). As a control, we verified that it could not rescue the SIL_{2all} mutant. Again, we interpret these results as indicating that the B and D isoforms are essential and the *kah93* partially reduces the function of the B isoform.

4.2.13 Gut and neuronal expression of NURF-1.D *is important for* C. elegans' *growth*

As a continuation of the rescue analysis, I also tried to understand if the expression of *NURF-1.D* in different cells will have a different effect. I put different cell-specific

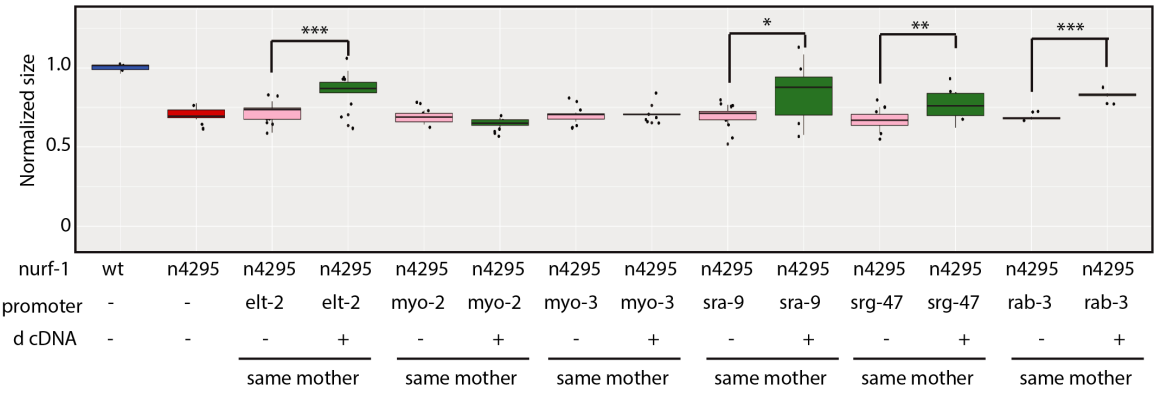

Figure 4.13. Transgenic tissue-specific promoter driven *nurf-1.d* cDNA rescue of *n4295* growth. Tissue-specific promoters include *elt-2*, express specifically in the gut; *myo-2*, express specifically in pharynx; *myo-3*, express specifically in body wall muscle; *sra-9*, express specifically in ASK sensory neurons; *srg-47*, express specifically in ASI amphid neurons and *rab-3*, express in all types of neurons.*** indicates a significant difference at p-value<0.01 level. ** indicates a significant difference at p-value<0.05 level. * indicates significant difference at p-value<0.1 level.

promoters before *nurf-1.d* cDNA sequence and tried to rescue the slow growth phenotype of classical allele *nurf-1(n4295)*. Promoters tested are *elt-2*, express specifically in the gut; *myo-2*, express specifically in pharynx; *myo-3*, express specifically in body wall muscle; *sra-9*, express specifically in ASK sensory neurons; *srg-47*, express specifically in ASI amphid neurons and *rab-3*, express in all types of neurons. *elt-2, sra-9, srg-47,* and *rab-3* driven *NURF-1.D* expression can partially rescue the slow growth phenotype (**Figure 4.13**), suggesting gut and neuronal expression, but not pharyngeal or body muscle expression of *NURF-1.D* is important for *C. elegans*'s growth.

4.2.14 Lack of nurf-1.d *is partially causative for ARL_{del} slow reproduction and growth*

Take advantage of the *nurf-1.d* cDNA genomic transgene, I backcrossed it to ARL_{del} to verify if the 60bp deletion impaired the function of *nurf-1.d* thus caused show reproduction

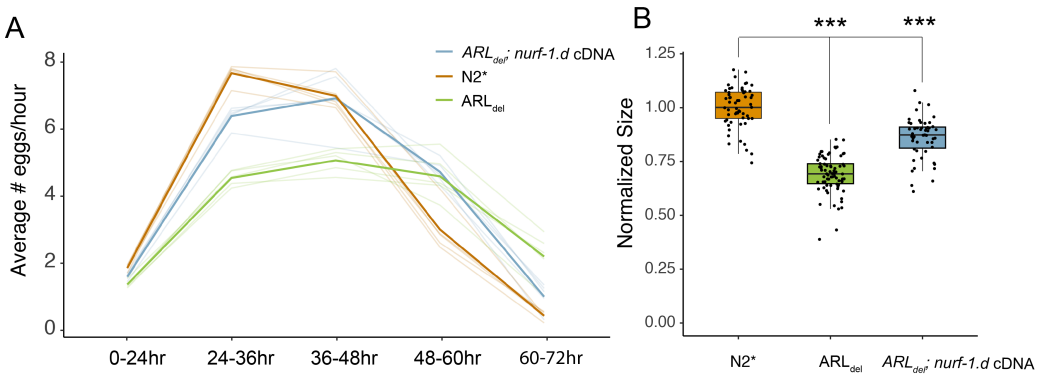

Figure 4.14. Integrated P*nurf-1.d::nurf-1* partially rescues of ARL$_{del}$ egg laying and fecundity.A) **)** Egg laying rate of N2*, ARL$_{del}$ and ARL$_{del}$ integrated *nurf-1.d* cDNA rescue. Integrated *nurf-1.d* cDNA rescue strain showed partially rescued later reproductive timing phenotype of ARL$_{del}$. B) Size of N2*, ARL$_{del}$ and ARL$_{del}$ integrated *nurf-1.d* cDNA rescue normalized to the average size of N2*. Integrated *nurf-1.d* cDNA rescue strain showed partially rescued slow growth phenotype of ARL$_{del}$.. *** indicates significant difference at p-value<0.01 level.

and growth phenotype. *nurf-1.d* cDNA transgene was able to partially rescue the slow growth and reproduction phenotype of ARL$_{del}$ (**Figure 4.14**). The full rescue was not observed may be due to the promoter region we use for *nurf-1.d* transgene was not sufficient for all related regulators to allow normal expression of *NURF-1.D*.

4.2.15 nurf-1.b and nurf-1.d function antagonistically during gametogenesis

The B and D isoforms are both required for reproduction, however, the mechanism that these isoforms operate through are could be different. Potentially both isoforms could participate as part of the NURF complex, cooperating together to regulate reproduction. However, the D isoform also could modify NURF activity by competing for binding with transcription factors or regions of the genome NURF is recruited to or act through a NURF-independent pathway. To gain insight into the molecular nature of the D isoform, we decided to test more carefully how the B and D isoforms regulated reproduction.

In order for hermaphrodites to produce a fertilized egg, gametogenesis must produce both male and female gametes. Initially, gametogenesis results in the production of sperm. After approximately 300 sperm is produced, a permanent sperm-to-oocyte switch occurs, and gametogenesis results in the production of oocytes until the animal dies or the gonad ceases to function. In *C. briggsae*, the B isoform is required for spermatogenesis in hermaphrodites. Animals lacking the B isoform bypass spermatogenesis and go straight into oogenesis, a phenotype known as the feminization of the germline (*fog*)[52]. We tested SIL_{2all} mutants, which lack the B isoform (**Figure 4.15A**), for the ability to produce sperm. Compared with N2 animals, which create ~300 sperm per animal, the number of sperm produced by SIL_{2all} animals was greatly reduced, reducing the production of only ~60 sperm (**Figure 4.15B**). These animals produced a normal number of oocytes, indicating that spermatogenesis seemed to be affected specifically as opposed to a general decrease in gametogenesis (**Figure 4.15C**).

We next tested SIL_{18} mutants, which lack the D isoform. These animals produced ~500 sperm, almost doubling the number of self-sperm normally produced (**Figure 4.15B**). These animals produced almost no oocytes (**Figure 4.15C**). These results suggest the D isoform is necessary for the switch from spermatogenesis-to-oogenesis. While animals that lack either the B or D isoform are unable to reproduce, the cause of sterility is different at the cellular level.

Finally, we performed similar experiments on SIL_{15} mutants, which also lack the D isoform but also have a truncated B isoform. These animals showed an intermediate phenotype,

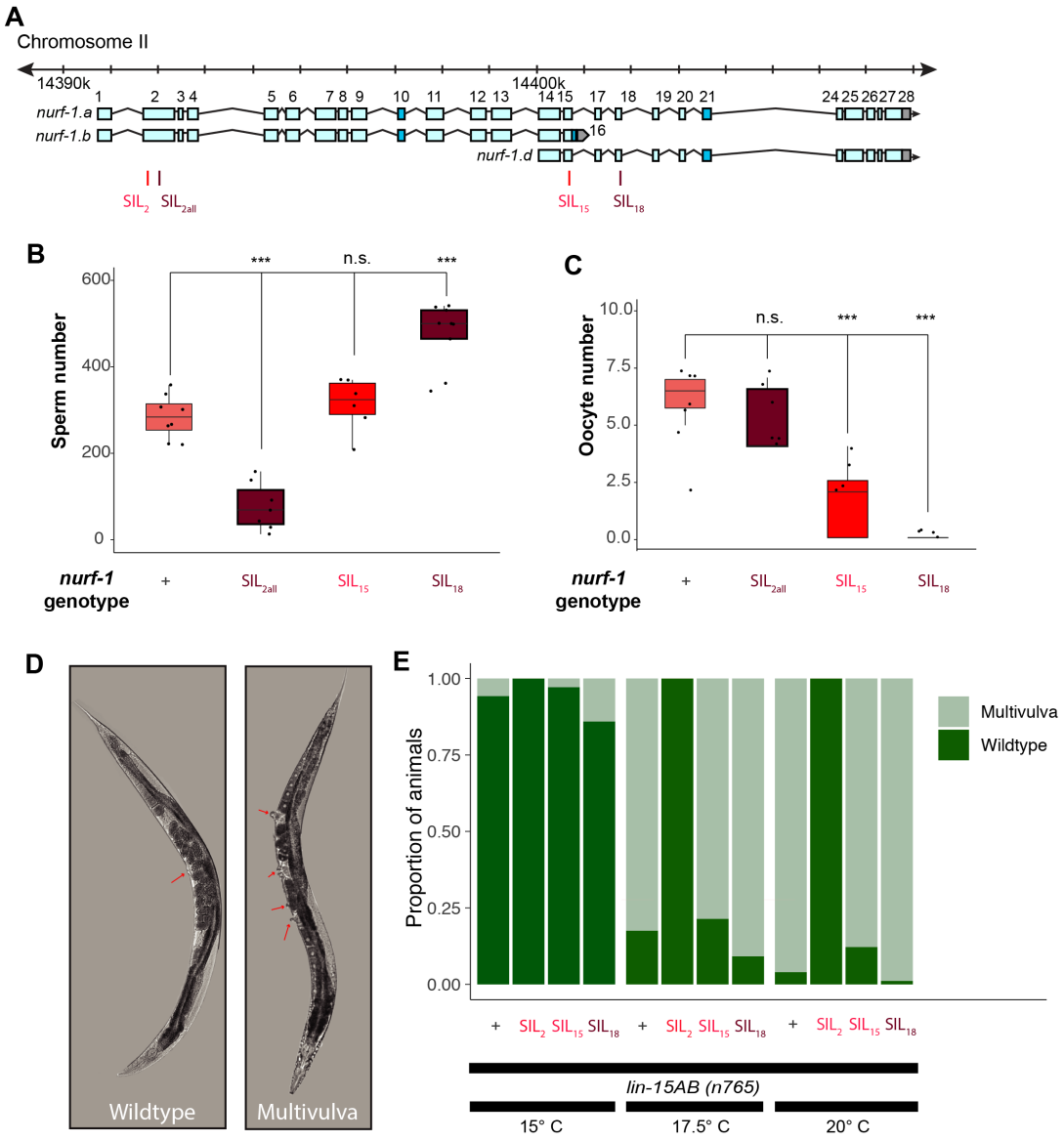

Figure 4.15. *nurf-1.b* and *nurf-1.d* function antagonistically.A) Relative positions of SIL_2, SIL_{2all}, SIL_{15} and SIL_{18} in *nurf-1*. **B)** Total sperm number of N2, SIL_{2all}, SIL_{15}, and SIL_{18} 12 hours post L4 stage. .*** indicates significant difference at p-value<0.01 level. n.s. indicates no significant difference at p-value<0.1 level. **C)** Total diakinesis and mature oocyte number of N2, SIL_{2all}, SIL_{15}, and SIL_{18} 12 hours post L4 stage. .*** indicates significant difference at p-value<0.01 level. n.s. indicates no significant difference at p-value<0.1 level. **D)** Wildtype vulva and multivulva phenotype. Red arrow is pointing to vulva position. **E)** The proportion of wildtype and multivulva animals for N2, SIL_2, SIL_{15}, and SIL_{18} in *lin-15AB(n765)* background at three different temperatures. *nurf-1(+)* animals have a significantly lower percentage of multivulva animals than SIL_{18} at 15°C and 17.5°C. *nurf-1(+)* animals have a significantly higher percentage of multivulva animals than SIL_2 at 15°C and 17.5°C. Two sample z-test was applied to test significance at p<0.05 level. SIL_{15} had intermediate synMuv phenotype between SIL_2 and SIL_{18}.

with a normal number of sperm but reduced number of oocytes (**Figure 4.15B**, **C**). While these animals lack the D isoform, potentially the reduced activity of the B isoform allows other factors to transition the animals to oogenesis, resulting in the milder phenotype of the SIL_{15} animals (**Figure 4.15B**).

4.2.16 nurf-1.b and nurf-1.d also have an opposite effect on vulva development

nurf-1 was originally identified in *C. elegans* as a suppressor of multivulval phenotypes of *lin-15AB (n765)* mutants (**Figure 4.15D**); the *n4293* allele suppressed multivulval phenotypes but the *n4295* did not. Based upon our work indicating the n4295 is not a loss-of-function of the D isoform, we were curious if the stronger *kah96* allele might demonstrate a role for the D isoform in vulval development. We validated that the *lin-15AB(n765)* mutant displayed a temperature-dependent multivulval phenotype and that it could be suppressed by the SIL_2 allele (**Figure 4.15E**). Interestingly, while the effect was subtle, *kah96* enhanced the multivulval phenotype of *lin-15AB(n765).* These results further support the model that the B and D isoforms can regulate the same phenotype in opposite directions.

4.2.17 C-terminal PHD domains and Bromodomain are not essential for C. elegans

The above data indicates that the two C-terminal PHD domains and the C-terminal bromodomain are not essential for the function of the D isoform (compare the SIL_{18} and *n4295* phenotypes in **Figure 4.10C**). This is surprising, as one of the hallmarks of BPTF in mammals is the importance of these domains in recognizing histone modification on specific nucleosomes and stabilizing interactions of NURF with targeted regions of the

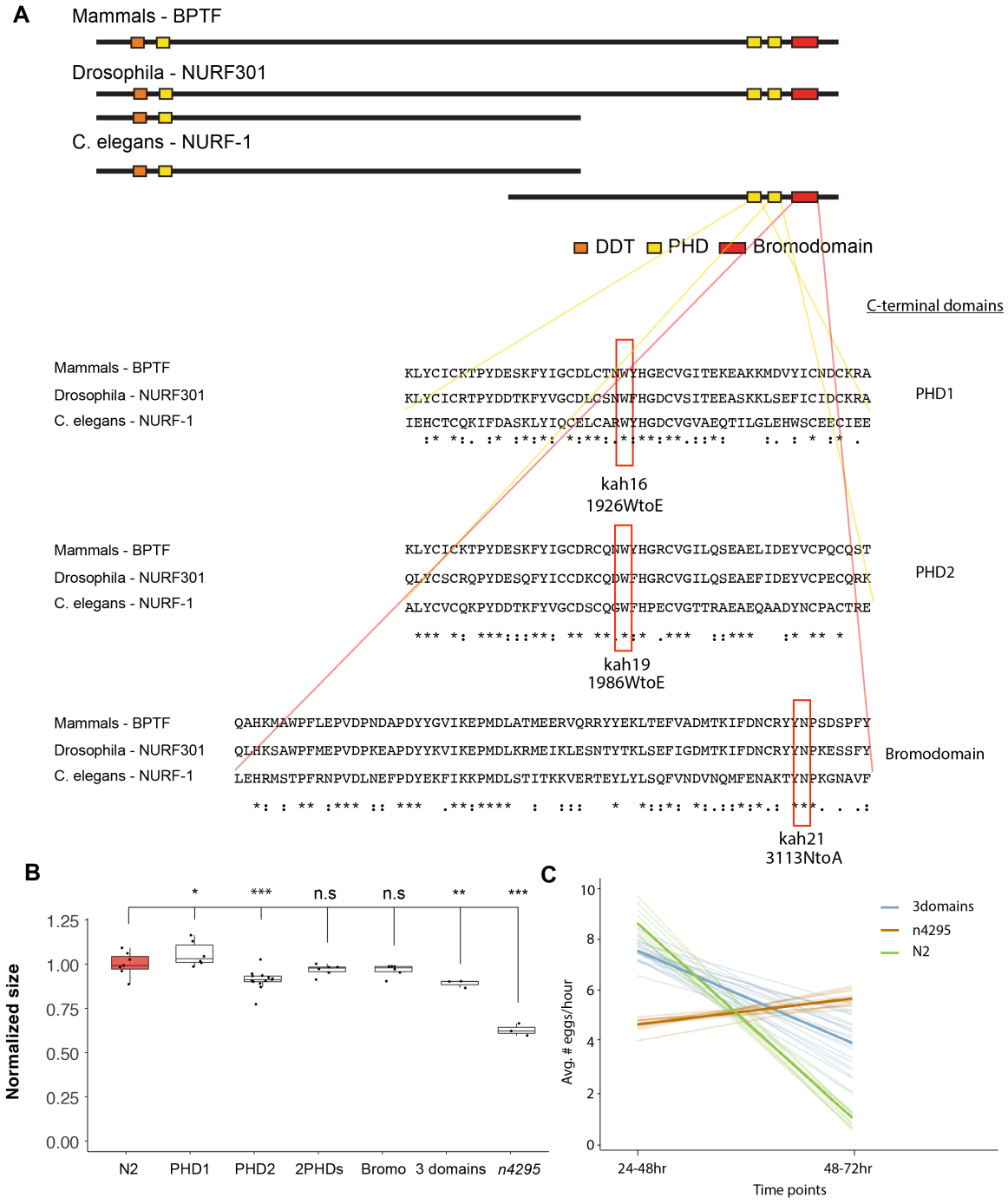

Figure 4.16. Histone marks recognition domains were not essential.A) Mutations to disrupt the function of different histone recognition domains are conserved across species. B) Size of N2, PHD domain mutants and Bromodomain mutants normalized to the average size of N2. Only the triple mutant that affects all three domains show a significant effect on growth, but this effect is smaller than the *n4295* mutation. *** indicates a significant difference at p-value<0.01 level. n.s. indicates no significant difference at p-value<0.1 level. C) The triple mutant affects all histone recognition domains affect egg laying rate,

but at a less significant level comparing to *n4295.* Egg laying rate was measured at two-time points when N2 and *n4295* show the biggest difference.

genome for remodeling chromatin state. For example, a single W to E substitution in the 2nd PHD domain resulting in loss of H3K4me3 recognition is unable to rescue BPTF function during Xenopus development [43]. Potentially, the importance of NURF-1 to recognize modified nucleosomes might have been reduced in the nematode lineage. Since the deletion in *n4295* potentially affects protein stability, we decided to follow the targeted substitution approach to study the importance of nucleosome recognition in an *in vivo* context. The tryptophan residue required for H3K4me3 recognition is conserved in both C-terminal PHD domains of *nurf-1* (**Figure 4.16A**). Using CRISPR/Cas9, I edited each PHD domain individually or together, replacing the conserved tryptophan with a glutamine residue, following the approach of Wysocka et al. I also tested the importance of the bromodomain in recognizing the H4K16ac mark by substituting a conserved asparagine necessary for the recognition of acetylated H4 with an alanine[53]. I tested these strains for their growth rate and compared them to *n4295* as a control. Each of the strains carrying one or all of the domain substitutions have no or little effect on animals growth comparing to the difference of N2 and *n4295* animals (**Figure 4.16B**). I tested the three-domain substitution strain on the reproductive rate at on two-time points (**Figure 4.16C**). Again, this strain showed significant differences from N2 but was also less affected than the *n4295* mutants.

4.2.18 Similar short C-terminal isoforms may exist in nurf-1's mammalian ortholog BPTF

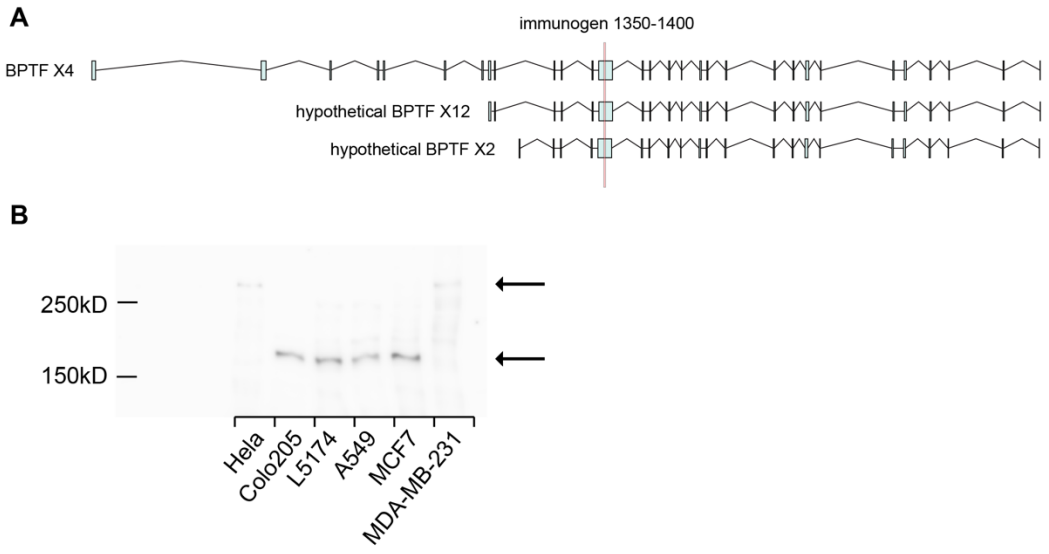

Figure 4.17. Short C-terminal BPTF may exist. A) Three representative forms of 21 predicted BPTF isoforms. An antibody targeting site at 1350-1400 amino acid region is shown in red. B) Western blot using the antibody targeting the 1350-1400 amino acid region of BPTF identified the full-length BPTF and a band between 150-250kD. Six human cancer cell lines were used: Hela, human cervical cancer cell line; Colo205 and L5174T, human colon cancer cell lines; A549, human lung cancer cell line; MCF7 and MDA-MB-231; human breast cancer cell lines. Long form BPTF was detected in Hela and MDA-MB-231 while the shorter reactive protein was detected in four other cancer cell lines.

It was inspiring to see the antagonistic function of different *nurf-1* isoforms, so I want to ask if it is also the case in mammals. Since there were also predicted short C-terminal BPTF transcripts from NCBI RefSeq platform[30], I decided to use western blot to verify the the existence of these isoforms. **Figure 4.17A** showed the representative long form BPTF X4 isoform, the two short predicted C-terminal isoforms X12 and X2. I looked up commercial antibodies that could target all three isoforms and found the Novus Biology anti-BPTF antibody NB100-41418 could target the 1350-1400 amino acid region of BPTF, which was shared by all predicted isoforms.

I performed western blot on six different human cell lines that came from the Georgia Tech community. I observed the ~322kD long form BPTF in cervical cancer cell line HELA and

breast cancer cell line MDA-MB-231 samples. From four out of the six different human cell lines, I observed a band between 150kD to 250kD molecular weight region (**Figure 4.17B**). But the predicted size of both short C-terminal BPTF transcripts was around 250kD. It was inspiring to see a consistently targeted band across different human cell lines, but this result could not provide enough evidence for the existence of new BPTF isoforms.

4.3 Methods

4.3.1 Strains

Near isogenic lines (NILs):

CX12311 (N2*) - kyIR1(V, CB4856>N2), qgIR1(X, CB4856>N2);

PTM66($NIL_{nurf\text{-}1}$) - kyIR87(II, LSJ2>N2); kyIR1(V, CB4856>N2), qgIR1(X, CB4856>N2);

CRISPR-generated allelic replacement lines (ARLs):

PTM88 (ARL_{del}): *kyIR1 (V, CB4856>N2); qgIR1 (X, CB4856>N2); nurf-1(kah3) II; spe-9(kah132) I*;

PTM416 ($NARL_{intron}$): *nurf-1(kah127) II*

Promoter-GFP lines:

PTM103 ($P_{nurf\text{-}1\_1st}$::gfp): *kahEx45;*

PTM104 ($P_{nurf\text{-}1\_1st}$::gfp): *kahEx46;*

PTM105 ($P_{nurf\text{-}1\_1st}$::gfp): *kahEx47;*

PTM106 ($P_{nurf\text{-}1\_14th}$::gfp): *kahEx48;*

PTM107 ($P_{nurf\text{-}1\_14th}$::gfp): *kahEx49;*

PTM108 ($P_{nurf\text{-}1\_14th}$::gfp): *kahEx50;*

PTM127 ($P_{nurf\text{-}1\_22nd}$::gfp): *kahEx52;*

PTM128 ($P_{nurf\text{-}1\_23rd}$::gfp): *kahEx53;*

PTM129 ($P_{nurf\text{-}1\_23rd}$::gfp): *kahEx54;*

PTM130 ($P_{nurf\text{-}1\_22nd}$::gfp): *kahEx55;*

PTM143 ($P_{nurf\text{-}1\_17th}$::gfp): *kahEx66;*

PTM144 ($P_{nurf\text{-}1\_17th}$::gfp): *kahEx67;*

PTM148 ($P_{nurf\text{-}1\_2nd}$::gfp): *kahEx71;*

PTM149 ($P_{nurf\text{-}1\_2nd}$::gfp): *kahEx72;*

PTM152 ($P_{nurf\text{-}1\_18th}$::gfp): *kahEx75;*

PTM153 ($P_{nurf\text{-}1\_18th}$::gfp): *kahEx76;*

PTM157 ($P_{nurf\text{-}1\_21st}$::gfp): *kahEx80;*

PTM158 ($P_{nurf\text{-}1\_21st}$::gfp): *kahEx81;*

PTM161 ($P_{nurf\text{-}1\_26th}$::gfp): *kahEx84;*

PTM162 ($P_{nurf\text{-}1\_26th}$::gfp): *kahEx85;*

<u>CRISPR-generated STOP codons replacement lines</u>:

PTM98 (SIL_{23}): *nurf-1(kah11)II*;

PTM174 (SIL_{23}): *nurf-1(kah39)II*;

PTM188 (SIL_{26}): *nurf-1(kah53)II*;

PTM203 (SIL_{26}): *nurf-1(kah68)II*;

PTM316 (SIL_{1}): *nurf-1(kah90)II/ oxTi924 II*;

PTM317 (SIL_{2}): *nurf-1(kah91)II/ oxTi924 II*;

PTM318 (SIL_{2}): *nurf-1(kah92)II/ oxTi924 II*;

PTM319 (SIL_{15}): *nurf-1(kah93)II/ oxTi924 II*;

PTM320 (SIL_{15}): *nurf-1(kah94)II/ oxTi924 II*;

PTM322 (SIL_{18}): *nurf-1(kah96)II/ oxTi924 II*;

PTM323 (SIL_{18}): *nurf-1(kah97)II/ oxTi924 II*;

PTM325 (SIL_{19}): *nurf-1(kah99)II/ oxTi924 II*;

PTM332 (SIL_{2all}): *nurf-1(kah106) II/ oxTi924 II*

CRISPR-generated domain replacement lines:

PTM113 (PHD1): *nurf-1(kah16)II*;

PTM116 (PHD2): *nurf-1(kah19)II*;

PTM117 (PHD2): *nurf-1(kah20)II*;

PTM118 (Bromodomain): *nurf-1(kah21)II*;

PTM167 (Bromodomain): *nurf-1(kah32)II*;

PTM170 (double PHD): *nurf-1(kah20,35)II*;

PTM170 (double PHD): *nurf-1(kah19,36)II*;

PTM189 (3 domains): *nurf-1(kah19,36,54)II*;

PTM201 (PHD1): *nurf-1(kah66)II*;

PTM211 (double PHD): *nurf-1(kah66,kah73)II*

CRISPR-barcoded strain:

PTM229: *dpy-10(kah82)II*;

CRISPR-epitope tagged strain:

PTM420 (HAFLAG): *nurf-1(kah124,kah133) II;* PTM489 (*nurf-1.f* deletion; HAFLAG): *nurf-1(kah124,kah133, kah144) II*

MosSCI transgenic strains:

PTM337 (*nurf-1.d* cDNA_Mos): *kahSi7II;*

PTM517 (PTM88_dMos): *kyIR1 (V, CB4856>N2); qgIR1 (X, CB4856>N2); nurf-1(kah3) II; kahSi7 II;*

synMuv double mutants:

PTM342 (STOP2;lin-15AB): *nurf-1(kah91) / oxTi721 II; lin-15AB(n765) X*;

PTM343 (STOP2;lin-15AB): *nurf-1(kah92) / oxTi721 II; lin-15AB(n765) X*;

PTM344 (STOP15;lin-15AB): *nurf-1(kah93) / oxTi721 II; lin-15AB(n765) X*;

PTM345 (STOP15;lin-15AB): *nurf-1(kah94) / oxTi721 II; lin-15AB(n765) X*;

PTM346 (STOP18;lin-15AB): *nurf-1(kah96) / oxTi721 II; lin-15AB(n765) X*;

PTM347 (STOP18;lin-15AB): *nurf-1(kah97) / oxTi721 II; lin-15AB(n765) X*;

PTM348 (STOP1;lin-15AB): *nurf-1(kah90) / oxTi721 II; lin-15AB(n765) X*

Other double mutants:

PTM354(n4295;STOP18): *nurf-1(n4295, kah113) II/ oxTi924 II*;

MosSCI rescue strains:

PTM370 (STOP2;dMos): *nurf-1(kah91) II/ oxTi721 II; kahSi7*;

PTM371 (STOP15;dMos): *nurf-1(kah93) II/ oxTi721 II; kahSi7*;

PTM372 (STOP18;dMos): *nurf-1(kah96) II/ oxTi721 II; kahSi7*;

PTM373 (STOP2all;dMos): *nurf-1(kah99) II/ oxTi721 II; kahSi7*;

PTM376 (n4295;dMos): *nurf-1(n4295) II; kahSi7*

4.3.2 Promoter GFP plasmid construction

To retain specificity, 500bp upstream each predicted start codon was define as a promoter region. Different *nurf-1* promoter region were amplified through promoter specific primers carrying FseI and AscI restriction sites. Digestion followed by ligation to pPM11 (pSM-Psrg-36-gfp) vector carrying the coding region of GFP. The constructed plasmids were

then injected to N2 and progenies carrying stable extrachromosomal array were maintained for observation.

4.3.3 CRISPR STOP strains and DOMAIN strains construction

I generated the STOP Isogenic Lines (SIL) as described in 2.4.1 except using sgRNAs targeting different exon of nurf-1 gene and repair oligos with a targeted stop codon and restriction digest sites. For DOMAIN strains, I replaced conserved PHD and bromodomain amino acids per suggestion from Alexander Ruthenberg, Associate Professor at the University of Chicago who is specialized in Chromatin biochemistry and structure. Those replacements will kill the H3K4me3 binding for PHD domains[43] and H4K16Ac binding to bromodomain[53].

The sgRNA sequences and repair oligos are listed below:

target	***sgRNA+pam sequence (5' to 3')***	***repair oligo (5' to 3')***	***genotyping primer f (5' to 3')***	***genotyping primer r (5' to 3')***
dpy-10(cn64)	CTACCATAGGCACCACGAG(CGG)	CACTTGAACTTCAATACGGCAAGATGAGAATGACTGGAAACCGTACCGCATGCGGTGCCTATGGTAGCGGAGCTTCACATGGCTTCAGACCAACAGCCTAT	GTCAGATGATCTACCGGTGTGTCAC	GTCTCTCCTGGTGCTCCGTCTTCAC

target	*sgRNA+pam sequence (5′ to 3′)*	*repair oligo (5′ to 3′)*	*genotyping primer f (5′ to 3′)*	*genotyping primer r (5′ to 3′)*
nurf-1 exon1	TTTCGCTTTGAACGGCCACG(TGG)	TTTCGAAAAGTTTCGAAATTTTTTCAGATGGCTCCCAGCTGAGGCCGTTCAAAGCGAAAGCATCCGTCTGAATCCGG	GCAGGCCGGCCGTGCGCCTTTAGGGTGTG	CGACGTGGTGGAGTTGGATT
nurf-1 exon2	ATTTGCGACGTGGTGGAGT(TGG)	CCATCAATTTTCAGCCGACCCAAAAAAAATCCAACCAGCTGACGTCGCAAATCGCTAAAACGCCAAGAAGAAGACAT	GCAGGCCGGCCTTCGCGCCTGGGTAATACC	CGGCAGTTTTCGTCGTTCTG
nurf-1 exon2all	ATGAGGATCAGGTTGTTCAGG(AGG)	GACGATGAGTTCATGCTAAATGAGGATCAGGTTGTcagctgaGAAGAGGAAGAGTTGAATTTAACAGATATTAAAAT	GCAGGCCGGCCTTCGCGCCTGGGTAATACC	CGGCAGTTTTCGTCGTTCTG
nurf-1 exon7	CTTCCTAACCCGGAAAAAGG(CGG) and CCTAACCCGGAAAAAGGCGG(CGG)	TGAGCGAATTTTGACGCCATATACTTCTTTTTATCtcagctgTTTTTCCGGGTTAGGAAGCTCTCAGACATCTCGTT	ACTAACTGGAATGATCGAAGAAGA	CCGGTTCGAAACGCTTTCAG

target	*sgRNA+pam sequence (5′ to 3′)*	*repair oligo (5′ to 3′)*	*genotyping primer f (5′ to 3′)*	*genotyping primer r (5′ to 3′)*
nurf-1 exon15	ATCAAGGGACACCACCAC C(AGG)	AATTAGTTTTCTAATTTTC AGGAATATATGGATGATC AAGGGACACCACCCAGCT GACAACAAGTTCGCTACG TGCTTCAAGGCGGAAACT CCGGTACACCAAATGT	ACGCCAG TTGTATA CGGCAA	GATCGAG GACATTCG AAAACCC G
nurf-1 exon18	AATTCGAAT GGAAGGAAT GT(GGG)	CCTCATTTATCTCTCGTTC CTCTTCTTCGTTGAATTCG AATGGAAGGAATCAGCT GTTTGTCTGCATCCTGAA CAGAAAAAACTTAATAA AACGCATATTGTGATT	TTTTGCG TCGAGGC CCATAT	TCGAAAA ACGACACC CTTGA
nurf-1 exon19	AGAAGATTG GAGACATAT GG(TGG)	AAATCCGATTCGTTCCAT CACGTCTTCAGGTGACAgc tgACGAGCcagctgATATGTca gctgaCTTCTACCGT and TTCAGGTGACAgctgACGA GCcagctgATATGTcagctgaCT TCTACCGTCTTCCAACGA TTCGCCTGATGATAA	AAAAAGT CTCCAGT TTATTTTC CTACT	TCTCAAAA TGCCAAAA ACAACAA A

target	*sgRNA+pam sequence (5′ to 3′)*	*repair oligo (5′ to 3′)*	*genotyping primer f (5′ to 3′)*	*genotyping primer r (5′ to 3′)*
nurf-1 exon23	CAGTAGGCAGAATAACAGTT(CGG)	TGGGCAATATTGTGATGGTGCAGTGTAGGTTGCAGTAGGCAGAATAACAGCTGGTCAAACCATTTTTTCAAATTTTCAGATGGATAGGATCTAGAAATCTAGTTGGG	CACATCTGCGCGTATTCAGC	TTATCAGACGACGTGCCAGG
nurf-1 exon26	TGTGCTCATACGGTGTTCG(AGG)	TTTGAGCTCCTCGAGCTTCTCCTCGAACACCGTATCAGCTGACCATTCAGGAACCCAGTAGATTTGAATGAGTTCCC	GACAGCTGTCAAGGGTGGTT	CAGCATTGCCCTTGGGATTG
nurf-1 c-terminal phd1	CTGTGCGCTCGTTGGTACCA(TGG)	CGAAGCTGTATATTCAGTGCGAGCTGTGCGCTCGTgagTACCATGGCGACTGTGTTGGAGTCGCCGAGCAAAC	GAGACCTGAAATTTTTCCAG	ctcGGCGCGCCACGGTGTTCGAGGAGAAGCT
nurf-1 c-terminal phd2	GTGTGACAGCTGTCAAGGG(TGG)	GGTAAATTTGCAGATTTTACGTTGGGTGTGACtccTGTCAAGGGgagTTTCACCCGGAATGCGTTGGGACAACACGTGCC	GAGACCTGAAATTTTTCCAG	ctcGGCGCGCCACGGTGTTCGAGGAGAAGCT

target	*sgRNA+pam sequence (5′ to 3′)*	*repair oligo (5′ to 3′)*	*genotyping primer f (5′ to 3′)*	*genotyping primer r (5′ to 3′)*
nurf-1 c-terminal bromodomain	AGCATTGCCCTTGGGATTGT(AGG)	AACGACGTGAATCAAATGTTCGAAAATGCAAAAACgTACgctCCCAAGGGCAATGCTGTATTCAAGTGCGCCGAAA	ATTACAACTGCCCGGCAT	gagacagcTCAAAAATCATAAAGTCTCATAAGATTCGC
nurf-1 middle ha tag	TGGCACTTGCTCAGTTGTGG(TGG)	TTTTGTCAAATTTGGAGCCGTTTGGGGAACCTCTAggcgtagtcggggacgtcgtatgggtatcctcctcctcctcctcccTGcTGtTCgTCTGGgACcTGCTCgGTTGTaGTaGAAACTGCGAAACCAGTCGCGTCATCAGGCATGTC	ACGCCAGTTGTATACGGCAA	GATCGAGGACATTCGAAAACCCG
nurf-1c-terminal flag tag	RNA	TCGTTGGATATCGATTCGGATCAGCTGTTGCCACTcGAcGCcAAcCTcATGcgtCTcTAcGAcTTcGGAGCCGGATCTGATTATAAAGACGATGACGATAAG	CGACAAAAAGTTGATAGACG	CCGGTCTCGACACAATTTTT

target	*sgRNA+pam sequence (5′ to 3′)*	*repair oligo (5′ to 3′)*	*genotyping primer f (5′ to 3′)*	*genotyping primer r (5′ to 3′)*
		CGTGACTACAAGGACG ACGACGACAAGCGTGA TTACAAGGATGACGAT GACAAGAGATGAGATT TTTCCTCTTGTTTTTTTA TGTTTTTCGT		
nurf-1 exon23 deletion	GTCTCTCGT CAAAAGGT GGG(CGG) and TATTAGACG ACTCTAAGA GG(CGG)	N/A	gcaGGCCG GCCGCTC GATACGT TAAAATT TGAACT	CGCGTCGA AAATCTTC TGACA
dpy-10(kah82)	CTACCATAG GCACCACG AG(CGG)	cacttgaacttcaatacggcaagatga gaatgactggaaaccgtactgctcgtg gtgcctatggtagcggagcttcacatg gcttcagaccaacagcctat	GTCAGAT GATCTAC CGGTGTG TCAC	GTCTCTCC TGGTGCTC CGTCTTCA C

4.3.4 Egg laying rate analysis for STOP strains

Egg laying rate experimental plates were prepared the same as in 2.4.3. For strains that have severely reduced fertility when homozygote, one-fourth larval stage (L4) nematode was transferred to the 3 cm experimental plate seeded with 50µl OP50. The number of eggs laid was measured every 12 or 24 hours, and eggs laid per hour was calculated by dividing the time range and the number of animals left on each plate at each time point. At least 10 replicates were assayed for each strain.

For other strains, six fourth larval stage (L4) nematode was transferred to the 50µl experimental plate. The number of eggs laid was measured every 12 or 24 hours, and eggs laid per hour was calculated by dividing the time range and the number of animals left on each plate at each time point. Six replicates were assayed for each strain.

Fecundity was calculated by summing up all eggs laid for each worm.

4.3.5 Growth analysis for STOP strains

Due to heterozygosity of multiple SILs, at 72 hours, only animals with no fluorescence were recorded for one minute using a Videomach camera. Other steps are the same as described in 2.3.3.

4.3.6 Compound heterozygotes egg laying analysis

Compound heterozygote strains were made just before any growth or egg-laying assay. Two strains, strain 1 and strain2, were first decided to make compound heterozygotes. If strain 1 was balanced with GFP marker, then it was first crossed with a homozygote RFP balancer males. The F1 RFP males without green fluorescence would carry the strain 1 mutation and this strain was crossed with strain 2 hermaphrodites. F2 non-fluorescent worms would carry mutations from both strains. If strain 1 was balanced with RFP marker,

then homozygote GFP marker males were used for the first cross, and F1 GFP males without red fluorescence would proceed for the next cross.

4.3.7 Compound heterozygotes growth analysis

Compound heterozygotes worms were prepared by crossing two heterozygotes SILs carrying different mutations with GFP or RFP balancing markers overnight and allow the hermaphrodite worm to lay eggs for 2 hours. At 48 hours, non-fluorescent animals with similar size were picked to the same plate. At 72 hours, animals from the same plate were recorded for 1 minute to decide their size. All animals were genotyped to confirm their heterozygosity after assay.

4.3.8 heat shock RNAseq samples preparation

20 N2 and PTM489 gravid adult hermaphrodites were picked to fresh 6cm NGM agar plates and let lay eggs for 3 hours. Worms were then picked off and the eggs remained on plates were cultured at 20°C until they reach L4 stage. Heat shock assay plates were then wrapped with parafilm and placed in a water bath preheated to 34°C for 2 hours and 4 hours. Worms were either collected right after heat shock or 30 minutes later for the recovery group. Each strain and condition had 3 independent replicates on the same day.

4.3.9 Transcript differential expression analysis

RNAseq and transcriptome analysis were performed as described in 3.3.3.

Kallisto program was used to quantify abundances of *nurf-1* transcripts[49]. We first created our own reference transcriptome by modifying the transcripts in Wormbase[45] published reference transcriptome so that it only has *nurf-1. a nurf-1.b, nurf-1.d, nurf-1.f*

and *nurf-1.q* isoforms. Afterward, we followed the Kallisto pipeline and quantified transcript abundance for each *nurf-1* isoform.

4.3.10 Sperm counting analysis

4 N2, PTM332, PTM319, and PTM332 gravid hermaphrodites were picked to fresh 5.5cm NGM agar plates. After 3 days, 20-30 L4 worms were picked to a new NGM plate and let grow at 20°C for 12 hours. Worms were then picked to a drop of M9 buffer on a Fisher Superfrost Plus slide (22-037-246). Fixation was done through applying 95% ethanol for three times. A drop of DAPI solution with Vector Laboratories Vectashield Mounting Medium with DAPI (H-1500) was added and the coverslip was sealed with nail polish. Z-stack images were captured through Olympus microscope (Olympus xx) under 40x objective through DIC and DAPI channels. Diakinesis and mature oocytes were counted while imaging and sperm number were counted manually on ImageJ through CellCounter plugin.

4.3.11 Promoter nurf-1.d cDNA extrachromosomal array rescue assay

<u>Promoter cDNA GFP plasmid construction</u>

pSM - *Pnurf-1.d::*nurf-1.d-SL2-GFP

First, nurf-1.d cDNA was reverse transcribed from C. elegans RNA and amplified using primers contains NheI restriction sites, this cDNA piece was then digested and ligated to a pSM vector. Second, a 2890bp long promoter region right before nurf-1.d isoform was amplified with a forward primer including FseI and a reverse primer including AscI restriction sites. This promoter region was then digested and ligate to the vector

constructed in step 1. Third, an SL2-GFP sequence from pPM7 (pSM-SL2-GFP) was cut and ligated to the new vector using KpnI and SpeI restriction sites.

pSM – *Ptissue-specific-promoter::*nurf-1.d-SL2-GFP

For different promoters, specific promoter region was amplified from N2 genomic DNA with FseI in the forward primer and AscI in the reverse primer. Restriction digest followed by ligation to the pSM - *Pnurf-1.d::*nurf-1.d-SL2-GFP vector would place different tissue-specific promoters in the correct region.

Promoter cDNA GFP growth analysis

Promoter cDNA GFP plasmids were injected into *n4295* worms and progenies carrying fluorescent extra-chromosomal array will be selected for growth assay. 20 fluorescent gravid adults were picked to assay plates for egg laying for two hours. After 72 hours, fluorescent worms and non-fluorescent worms were imaged separately to analyze the effect of *nurf-1.d* cDNA extrachromosomal array rescue.

4.3.12 nurf-1.d cDNA MosSCI rescue

MosSCI vectors construction

pCFJ151 - *Pnurf-1.d::*nurf-1.d-SL2-GFP:

Using the pSM - *Pnurf-1.d::*nurf-1.d-SL2-GFP plasmid constructed during *nurf-1.d* cDNA rescue assay, I first inserted a sequence containing FseI and AscI restriction sites into the pCFJ151 vector using NEB Q5 site-directed mutagenesis kit. Then the Pnurf-1.d::nurf-1.d-SL2-GFP cassette was inserted into the modified pCFJ151 backbone through FseI and SpeI digestion and ligation.

pCFJ151 - Pnurf-1.d::nurf-1.d-FLAG-SL2-GFP:

The procedure was mostly the same as pCFJ151 - Pnurf-1.d::nurf-1.d-SL2-GFP construction except that before inserting the Pnurf-1.d::nurf-1.d-SL2-GFP cassette to the modified pCFJ151 backbone, we used NEB Q5 site-directed mutagenesis kit and inserted a spacer and three FLAG sequence just before *nurf-1.d* stop codon.

MosSCI- generated cDNA transgenic lines

MosSCI strain construction was done following standard protocol from Frøkjær-Jensen et. al[54]. Injection mix was prepared as following: 38ng/ul pCFJ601 (Mos1 transposase), 30ng/ul pCFJ151 - P*nurf-1.d::nurf-1.d-SL2-GFP* or pCFJ151 - P*nurf-1.d::nurf-1.d-FLAG-SL2-GFP* (insertion vector with homologous arms), 2.5ng/ul pCFJ90 (*Pmyo-2*::mCherry (pharynx muscle)) 5ng/ul pCFJ104 (*Pmyo-3*::mCherry (body muscle)). This injection mix was injected into EG6699 uncoordinated animals. Three injected animals were placed on a single plate at 30 ^{o}C to facilitate starvation. After 5 days, single normally moving animal with GFP fluorescence and no red fluorescence were singled to new NGM plates and let proliferate. The progenies were then singled and only homozygotes without uncoordinated progenies were maintained. The homozygotes were then backcrossed to N2 for 4 generations to get rid of *unc-119(ed3) III.*

nurf-1.d cDNA MosSCI rescue experiment

The strain carrying *nurf-1.d* cDNA is PTM337. This strain was crossed with multiple *nurf-1* mutant strains to perform rescue experiment. For mutants that have homozygote deficiency (SIL_{2all}, $SIL_{15,}$ and SIL_{18}), double mutants were kept as heterozygotes at mutation site and homozygotes at *nurf-1.d* cDNA site. Egg laying analysis and growth analysis were performed as described in 4.3.4 and 4.3.5.

For *n4295* and ARL_{del} rescue analysis, homozygote animals with rescuing *nurf-1.d* cDNA transgene was assayed.

4.3.13 synMuv scoring

6 HA353 fluorescent L4s and heterozygote other test strains were picked to fresh 5.5cm NGM plates and grow at 3 different temperatures: 15°C, 17.5°C and 20°C. Until most worms reach L4 stage, about 100 F1 non-fluorescent L4 animals were then picked to new plates and let grow at the temperature they previously were cultured in. Worms were scored under a dissecting microscope at adulthood.

4.3.14 Western blots

4 N2, PTM420 and PTM 352 gravid hermaphrodites were picked to fresh 5.5cm NGM agar plates. Worms were collected just prior to starvation using M9 buffer and stored at -80°C until protein extraction. Protein sample preparation is as followed:

At least 4 plates of worms were used for protein isolation. Worms were condensed by centrifugation and 2x sample buffer (62.5 mM Tris-HCl pH 6.8M, 2.5 % SDS, 0.002 % Bromophenol Blue, 0.7135 M (5%) β-mercaptoethanol10 % glycerol) was added in 1:1 w/v ratio. 1ul of 500mM EDTA and 1ul of Halt protease inhibitor cocktail (100x) (Catalog number: 78430) were added for every 100ng of worm sample. We vortex the protein sample for 90 seconds and let sit on ice for about 1 minute. Samples were then sonicated in a Bransonic 0.5 gallon ultrasonic bath filled with hot water > 80°C for 10 minutes and immediately placed on ice for 2 minutes. We boil the samples for 5 minutes and placed on ice for cooling down. We then centrifuge the sample at 12,000rpm for 5 minutes and transfer the supernate to new tubes and finish protein lysate preparation.

All samples were loaded on 5% SDS-PAGE gel at 3ul, 5ul and 7ul volumes followed by Coomassie blue staining and washing steps. Gels were then dried using DryEase Mini-Gel Drying System (Invitrogen, Catalog number: NI2387). We then determine the loading volume for different samples to ensure consistent protein loading.

We then loaded each sample with defined volume to either a freshly made 6% or 10% SDS-PAGE gel and run at 25mA. Gel samples were then transferred in 10mM CAPS pH10.5 buffer at 20V and 20mA for 17hrs to a PVDF membrane. Protein products with HA tag were detected using 1:500 anti-HA antibody (Life Technologies, Catalog number: 326700) and protein products with FLAG tag were detected using 1:1000 PIERCE ANTI-DYKDDDDK antibody (Life Technologies, Catalog number: MA191878).

4.3.15 Nanopore sequencing data analysis

Nanopore sequencing data was shared from Timp lab at Johns Hopkins University. They shared the bam file reads in *nurf-1* gene region. I cleaned the duplicated reads and analyzed them through IGV genome viewer.

CHAPTER 5. Discussion

5.1 Conclusions

5.1.1 Laboratory evolution repeatedly target nurf-1

My thesis studies in chapter 2 and chapter 3 demonstrated that beneficial alleles in *nurf-1* were fixed in two laboratory strains of *C. elegans*. While parallel evolution is often observed in response to environmental shifts, there is a substantial difference in the growth conditions of LSJ2 and N2. LSJ2 was raised in liquid, axenic media using beef liver and soy peptone extract as a food source, which is an unnatural, poor quality food source for *C. elegans* nematodes. Wild *C. elegans* usually feed on rotten fruits with complex bacteria composites. LSJ2 animals spent approximately one month growing in this unnatural media until starvation and then were transferred to a fresh liquid culture. During the one month time, a lot of bad things accumulate: dead worms, wastes, pheromones, etc.. The major concern for LSJ2 animals is survival. N2 animals were cultured on agar plates seeded with *E. coli* OP50. Although *E. coli* is not a natural food source for them, they can easily consume and metabolize the bacteria. Animals were transferred every three days to avoid starvation. In this condition, survival is not a primary concern; each animal can eat as much food as possible and bare as many offspring as they can before transferring. The transfer process is either by picking 6 adult animals or chunk a small population to new plates. The major concern is no longer survival but reproduction.

The discrete shift in environment resulted in the evolution of multiple traits in N2 and LSJ2 animals, including reproductive timing, lifespan, survival, and growth rate[19]. These traits are all classical life history trade-off traits that are indicators for how an organism

distributes its resources. LSJ2 animals prioritize individual survival over reproductive speed, as they live longer, grow slower, and are less affected by stressors. These changes may be linked to the unnatural and poor nutritious value of liquid axenic media [55]. A strain of the related species *C. briggsae* that grew in the same media evolved similar changes: slower growth rate and reproductive timing changes[56]. These observations are all consistent with the hypothesis that the growth environment in this media creates selective pressure on fitness proximal traits. The 60bp deletion in *nurf-1* affects 18 amino acids and 3'-UTR. It resulted in dramatic phenotypic change as we observed in ARL_{del}. As far as I know, it was the first time a complex variation was confirmed by reversing it back to the ancestral state through CRISPR/Cas9 genome engineering. To understand how the 60bp deletion affects the function of *nurf-1* gene. I first looked at RNA sequencing reads aligned to *C. elegans* transcriptome and observed no difference in the quantity of *nurf-1* RNA except for last 18 amino acid change. My hypothesis is that the 60bp deletion will either affect *NURF-1* function or where it expresses. By rescue assay using integrated *nurf-1.d* transgene utilizing a universal *unc-54* 3'-UTR, I verified that *nurf-1.d* isoform was affected by this deletion but was not the only change.

To further understand how this 60bp deletion works and why this 60bp deletion was selected during evolution, I propose more experiments should be done. To test the 60bp deletion's effect on the 3'UTR, we can use previously constructed promoter-GFP vector, and perform genetic manipulations to change its 3'-UTR to both versions of *nurf-1* 3'-UTR and compare the differential expression of *nurf-1* by observing the expression pattern of GFP fluorescence. To study the effect of loss of 18 amino acids, we can use

CRISPR/Cas9 genome editing system to delete the last 18 amino acids coding sequences without affecting the 3'-UTR and test if there exist phenotypic changes.

N2 animals prioritize reproduction over survival since they live shorter, grow faster, and less resistant to stress but lay eggs early and fast. The fitness effect of *nurf-1* intron SNP on agar plates was a little unexpected although we have evidence that there exists another variation in the chromosome II QTL that affects the same set of genes in the opposite direction. From the fitness proximal traits I measured, only reproduction showed a small significant difference ($p<0.1$). The fitness advantage of N2 evolved *nurf-1* intron SNP is subtle in presence of *npr-1* and *glb-5* derived allele. However, it is also interesting to see how this intron SNP interacts with these two derived alleles that have been reported to dramatically affect animal's fitness[27]. I have made the ARL_{intron} strain when creating $NARL_{intron}$. From data not shown here, there' also a significant difference in fitness for ARL_{intron} and N2*. Unfortunately, I was not able to observe a significant difference in reproduction for ARL_{intron} and N2*. It is reasonable that the effect of *nurf-1* intron SNP is hard to follow due to its unpredictable effect on genes. This intron SNP locates in the middle of the 2nd intron of *nurf-1*. It's neither at splice junctions nor at splice branchpoint site. However, multiple studies found intron SNPs that far away from splice sites altering transcription activities[57], [58]. I then inspected the RNA sequencing results of N2 and $NARL_{intron}$ and found no difference in *nurf-1* RNA expression. This reduced the possibility that this intron SNP might affect splicing site that results in the emergence of a new transcript or abnormal expression of a transcript. However, it is interesting that I caught a time point (52hour post egg laying) when there exist about 3000 differentially expressed genes between N2 and $NARL_{intron.}$ Further transcriptome analysis found these genes are

mostly molting related genes. It was unexpected to find so many genes involved in molting since my primary goal was to examine if there exist differential expression in spermatogenesis genes during active spermatogenesis. Unfortunately, I caught a timepoint when L4/adult molting was very active. But the differential expression of molting related genes suggests that N2 and $NARL_{intron}$ have a difference in development.

To study why this intron SNP was affected will require further examination. As hermaphrodites, *C. elegans* start spermatogenesis at L3/L4 molting stage and transit to oogenesis at L4/adult molting stage [59]. At my hand, worm reach mature L4 stage at 48 timepoints, and I sampled RNA from worms 4 hours post L4 stage when most animals were actually undergoing molting to the adult stage, I think a good start point is to measure the gene expression difference right after animals reached the L4 stage and is actively undergoing spermatogenesis. This can avoid observing molting related genes so that we can examine the expression differences of genes important for spermatogenesis. Furthermore, from the RNAseq results, I didn't observe difference in RNA for *nurf-1* gene, which makes the function of this intron SNP unknown.

With the emergence of genome-wide association studies, numerous intron variations were identified [60]. However, very little studies focus on intron SNPs. It is well-recognized that intron variations may affect a gene in multiple aspects - they may affect mRNA alternative splicing or change gene expression if they serve as enhancers, but intron SNPs seems to be less 'attractive'. Fundamental genetic studies usually chose to follow-up variations seem to be 'big' – non-synonymous mutations, insertion/deletions, etc.. My finding on this *nurf-1* intron SNP's fitness advantage in its evolving encourages researcher to abandon their prejudice on variations to study. I found this intron SNP did not affect either the splicing

donor/acceptor nor the splicing branch point. But this cannot rule out that this intron SNP affect splicing or serve as enhancer/repressor element for *nurf-1* or other locations nearby. Multiple other databases, like ATAC-seq, histone modification ChIP-seq may provide insight into how this intron SNP functions in *C. elegans.* Also, to test if this intron SNP work as regulator element in the promoter region of *nurf-1,* a promoter-driven GFP plasmid can be used to test this hypothesis.

5.1.2 nurf-1 *isoforms function antagonistically*

Organisms evolves under selective pressure. Cats lack sweet receptor *Tas1r2* due to its food choice[61], *Caenorhabditis* species evolved in liquid culture lacks chemoreceptors *srg-36* and *srg-37* due to its high phenorome living environment [3]. In these cases, strong selective pressure was put on one aspect of an animal's life history, resulting in alteration of genes downstream (sensory receptors) of the regulatory networks. However, researchers also found parallel evolution of *cis*-regulatory changes in input/output genes during development [62]. For example, the *cis*-regulatory change in *shavenbaby* causes dorsal hair loss in *Drosophila sechellia* and its related species [1]. My finding of the parallel evolution of *nurf-1* gene is unique. *nurf-1* encodes an ortholog of human BPTF, as a large-effect, pleiotropic regulator of many of the LSJ2 life-history changes. BPTF is a subunit of NURF, a chromatin-remodeling complex that modifies transcription and promotes proliferation and differentiation of a number of tissues in an organism-specific manner [43]. Unlike coding region alterations in downstream sensory receptors or *cis*-regulatory changes in transcription factors, *nurf-1* has both coding region change and potential *cis*-regulatory change in response to the different environment. But why this chromatin remodeling factor was selected twice during *C. elegans's* adaptation is uncertain.

To my knowledge, BPTF/NURF has not previously been described as a regulator of fitness, but clues can be found from prior researches. Orthologs of NURF-1 have been reported to regulate a large number of traits but there's a little clue to link these traits together. Regulation of fitness proximal traits could be a comprehensive way to consider NURF-1/BPTF function in other species based on its conservation throughout evolution. In *Drosophila* melanogaster, NURF301 regulates homeotic gene expression, stress response, male X chromosome morphology, pupation and reproduction [42]–[44]. In mammals, BPTF regulates neural development, immune response, morphological development and cancer progression [33], [34], [63]. These are all traits that can be fit into the context of fitness.

BPTF/NURF is a complex location. It has 16 predicted isoforms in *C. elegans*, 25 predicted isoforms in human and 3 verified isoforms in *Drosophila.* Studies in mammals only focused on the full-length isoforms and little attention was paid on the short predicted isoforms. Study on NURF301 included two major isoforms: the long form and a short isoform NURF301C that only contains the N-terminal part of the gene. Kown et. al. reported that NURF301C was required for gametogenesis and 20% of NURF301's target will be affected if this short isoform was removed[44]. Inspired by *drosophila* NURF301 research and Erik Andersen's research on *C. elegans*'s *nurf-1* [46]*,* I decided to study its isoform-specific function.

In *C. elegans*, the 16 *nurf-1* isoforms can be classified into 9 representative forms including the long forms that have all functional domains, short N-terminal forms lack histone marks recognition domains PHD domain and bromodomain, short C-terminal forms that lack the ISWI interacting domain DDT. To my surprise, two *nurf-1* isoforms actually work

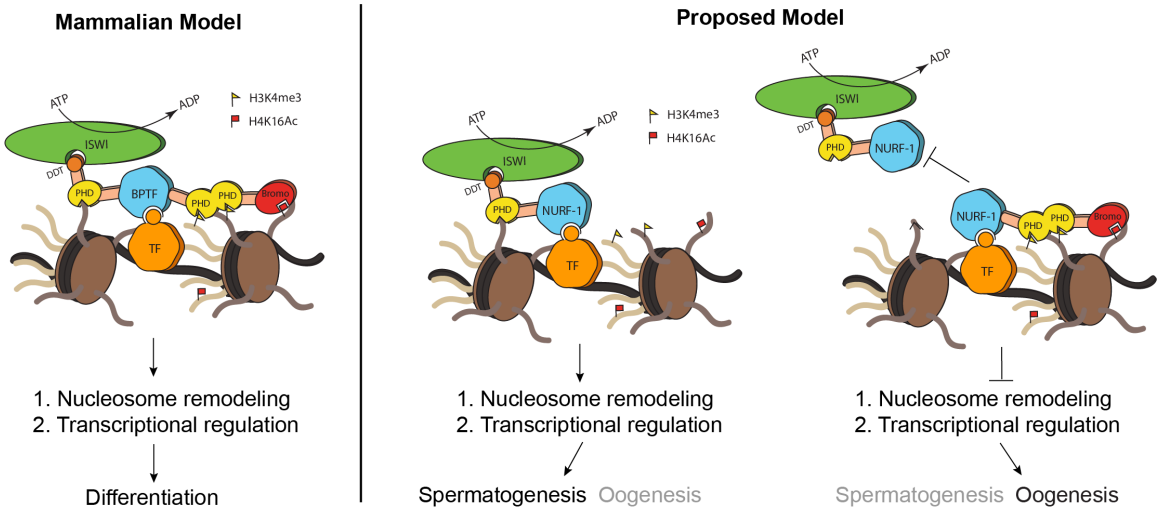

Figure 5.1. Proposed competing model for NURF-1 function. In mammals, only long-form BPTF is confirmed and well-studied. The proposed working model indicates how it interacts with ATPase ISWI to regulate differentiation. In *C. elegans*, we found the antagonistic effect of two *nurf-1* isoforms which match our proposed model that *NURF-1.B* regulates spermatogenesis and these two isoforms compete with each other to decide the timing of oogenesis.

antagonistically (here I refer as Yin-yang) while the full-length isoform seems to have no biological function in *C. elegans*. N-terminal *NURF-1B* regulates spermatogenesis while the C-terminal *NURF-1.D* regulates the timing of spermatogenesis to oogenesis transition. Both isoforms have a shared region which may interact with transcription factors. Our model for these two isoforms is that *NURF-1.B* work with a set of transcription factors during spermatogenesis when animals develop to the almost young adult stage, *NURF-1.D* express and work with the same set of transcription factors. Its histone mark recognition sites will help *NURF-1.D* outcompete *NURF-1.B* to initiate oogenesis (**Figure 5.1**). This finding is very interesting to us and updated how we usually think nucleosome remodeling factor functions.

However, more research could be useful to thoroughly study the function of this gene. My proposed antagonistic function of two *nurf-1* isoforms require further experimental support. From my model, *NURF-1.B* expresses at spermatogenesis stage while *NURF-1.D* expresses at spermatogenesis to oogenesis transition stage. This can be tested by analyzing the tissue- and time-specific expression of these two *nurf-1* isoforms, especially in the germline. I also proposed that these two isoforms work with similar transcription factors and compete with each other, this can be tested using immunoprecipitation-mass spectrometry (IP-MS) experiment take advantage of the epitope-tagged strains I constructed before. In support of my proposed model, I expect to see expression difference of *NURF-1.B* and *NURF-1.D* at different place or different stage. I also expect that the IP-MS experiment would find similar transcription factor targets for both isoforms.

The PHD and Bromodomains in *C. elegans* were found to be unnecessary, but previous studies reported that they were essential in mammalian system [43]. This contradiction suggest *nurf-1* may work differently in these two organisms. Considering the two short isoforms' antagonistic function together, I want to ask if the important function of two short *nurf-1* isoforms is unique to *C. elegans* or not. It is important to look at NURF-1's mammalian ortholog BPTF system and examine the short isoforms' existence. In collaboration with Francisco lab in Spain who works on mammalian BPTF, we are trying to find evidence of the small BPTF isoform through western blot and mass-spectrometry techniques. If the short isoforms do not apply to mammalian system, then an interesting question will be when does the splitting function of *nurf-1* and why. This can be tested by look at *nurf-1* orthologs in close *Caenorhabditis* species.

My finding of elevated expression of *nurf-1.f* isoform suggests there were about 200 genes that were differentially expressed, but I was not able to find out where should I put *nurf-1.f* in stress response pathway. Further study on those regulated genes should be helpful. Previous research on *nurf-1*'s orthologs also found it was important for stress response, so I think this *nurf-1.f* isoform's function also contributes to *nurf-1*'s role as a life history regulator.

My thesis study on this *nurf-1* gene is a comprehensive example of how to investigate a gene's evolutionary role and how to study a complex multi-isoform gene. Although there's a lot more can be done for this gene, my findings strongly encourage studying genes at isoform level.

5.2 Publications

Here's a list of publications related to this thesis:

Xu W, Long L, Zhao Y, McGrath PT (*in prep*) **An adaptive intron SNV regulates a NURF subunit gene expressing Yin and Yang isoforms with opposite effects on cell fate.**

Large, E.E., **Xu, W**., Zhao, Y., Brady, S.C., Long, L., Butcher, R.A., Andersen, E.C. and McGrath, P.T., 2016. **Selection on a subunit of the NURF chromatin remodeler modifies life history traits in a domesticated strain of Caenorhabditis elegans**. *PLoS genetics,* 12(7), p.e1006219.

Large, E.E., Padmanabhan, R., Watkins, K.L., Campbell, R.F., **Xu, W**. and McGrath, P.T., 2017. **Modeling of a negative feedback mechanism explains antagonistic pleiotropy in**

reproduction in domesticated Caenorhabditis elegans strains. *PLoS genetics*, 13(5), p.e1006769.

REFERENCES

[1] E. Sucena and D. L. Stern, "Divergence of larval morphology between Drosophila sechellia and its sibling species caused by cis-regulatory evolution of ovo/shaven-baby.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 97, no. 9, pp. 4530–4, Apr. 2000.

[2] P. F. Colosimo *et al.*, "Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles.," *Science*, vol. 307, no. 5717, pp. 1928–33, Mar. 2005.

[3] P. T. McGrath, Y. Xu, M. Ailion, J. L. Garrison, R. A. Butcher, and C. I. Bargmann, "Parallel evolution of domesticated Caenorhabditis species targets pheromone receptor genes," *Nature*, vol. 477, no. 7364, pp. 321–325, Aug. 2011.

[4] S. J. Gould, *Wonderful life : the Burgess Shale and the nature of history*. 1989.

[5] Z. D. Blount, R. E. Lenski, and J. B. Losos, "Contingency and determinism in evolution: Replaying life's tape.," *Science*, vol. 362, no. 6415, p. eaam5979, Nov. 2018.

[6] S. Venkataram *et al.*, "Development of a Comprehensive Genotype-to-Fitness Map of Adaptation-Driving Mutations in Yeast," *Cell*, 2016.

[7] M. M. Riehle, A. F. Bennett, and A. D. Long, "Genetic architecture of thermal

adaptation in Escherichia coli," *Proc. Natl. Acad. Sci.*, vol. 98, no. 2, pp. 525–530, Jan. 2001.

[8] T. E. Wood, J. M. Burke, and L. H. Rieseberg, "Parallel genotypic adaptation: When evolution repeats itself," in *Genetica*, 2005.

[9] U. Johanson, J. West, C. Lister, S. Michaels, R. Amasino, and C. Dean, "Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time.," *Science*, vol. 290, no. 5490, pp. 344–7, Oct. 2000.

[10] C. C. Steiner, J. N. Weber, and H. E. Hoekstra, "Adaptive Variation in Beach Mice Produced by Two Interacting Pigmentation Genes," *PLoS Biol.*, vol. 5, no. 9, p. e219, Aug. 2007.

[11] M. W. Nachman, H. E. Hoekstra, and S. L. D'Agostino, "The genetic basis of adaptive melanism in pocket mice.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 9, pp. 5268–73, Apr. 2003.

[12] C. R. Linnen, E. P. Kingsley, J. D. Jensen, and H. E. Hoekstra, "On the Origin and Spread of an Adaptive Allele in Deer Mice," *Science (80-.).*, vol. 325, no. 5944, pp. 1095–1098, Aug. 2009.

[13] N. Flanagan *et al.*, "Pleiotropic effects of the melanocortin 1 receptor (MC1R) gene on human pigmentation.," *Hum. Mol. Genet.*, vol. 9, no. 17, pp. 2531–7, Oct. 2000.

[14] S. E. Bush, S. M. Villa, J. C. Altuna, K. P. Johnson, M. D. Shapiro, and D. H. Clayton, “Host defense triggers rapid adaptive radiation in experimentally evolving parasites,” *Evol. Lett.*, Mar. 2019.

[15] E. C. Andersen *et al.*, “Chromosome-scale selective sweeps shape Caenorhabditis elegans genomic diversity,” *Nat. Genet.*, 2012.

[16] K. A. Swan, D. E. Curtis, K. B. McKusick, A. V. Voinov, F. A. Mapa, and M. R. Cancilla, “High-throughput gene mapping in Caenorhabditis elegans,” *Genome Res.*, 2002.

[17] O. Thompson *et al.*, “The million mutation project: A new approach to genetics in Caenorhabditis elegans,” *Genome Res.*, 2013.

[18] G. R. Abecasis *et al.*, “An integrated map of genetic variation from 1,092 human genomes.,” *Nature*, 2012.

[19] E. E. Large *et al.*, “Selection on a Subunit of the NURF Chromatin Remodeler Modifies Life History Traits in a Domesticated Strain of Caenorhabditis elegans,” *PLoS Genet.*, vol. 12, no. 7, pp. 1–19, 2016.

[20] C. elegans Sequencing Consortium, “Genome sequence of the nematode C. elegans: a platform for investigating biology.,” *Science*, vol. 282, no. 5396, pp. 2012–8, Dec. 1998.

[21] P. T. McGrath *et al.*, "Quantitative mapping of a digenic behavioral trait implicates globin variation in C. elegans sensory behaviors.," *Neuron*, vol. 61, no. 5, pp. 692–9, Mar. 2009.

[22] M. de Bono and C. I. Bargmann, "Natural Variation in a Neuropeptide Y Receptor Homolog Modifies Social Behavior and Food Response in C. elegans," *Cell*, vol. 94, no. 5, pp. 679–689, Sep. 1998.

[23] A. Persson, E. Gross, P. Laurent, K. E. Busch, H. Bretes, and M. De Bono, "Natural variation in a neural globin tunes oxygen sensing in wild Caenorhabditis elegans," *Nature*, 2009.

[24] E. C. Andersen, J. S. Bloom, J. P. Gerke, and L. Kruglyak, "A Variant in the Neuropeptide Receptor npr-1 is a Major Determinant of Caenorhabditis elegans Growth and Physiology," *PLoS Genet.*, 2014.

[25] J. A. Arribere, R. T. Bell, B. X. H. Fu, K. L. Artiles, P. S. Hartman, and A. Z. Fire, "Efficient Marker-Free Recovery of Custom Genetic Modifications with CRISPR/Cas9 in Caenorhabditis elegans," *Genetics*, vol. 198, no. 3, pp. 837–846, Nov. 2014.

[26] D. J. Dickinson, J. D. Ward, D. J. Reiner, and B. Goldstein, "Engineering the Caenorhabditis elegans genome using Cas9-triggered homologous recombination," *Nat. Methods*, vol. 10, no. 10, pp. 1028–1034, Oct. 2013.

[27] Y. Zhao *et al.*, "Changes to social feeding behaviors are not sufficient for fitness gains of the Caenorhabditis elegans N2 reference strain," *Elife*, 2018.

[28] S. G. Alkhatib and J. W. Landry, "The nucleosome remodeling factor.," *FEBS Lett.*, vol. 585, no. 20, pp. 3197–207, Oct. 2011.

[29] R. Bowser, A. Giambrone, and P. Davies, "FAC1, a novel gene identified with the monoclonal antibody Alz50, is developmentally regulated in human brain.," *Dev. Neurosci.*, vol. 17, no. 1, pp. 20–37, 1995.

[30] E. V Wasmuth and C. D. Lima, "GeneBank," *Nucleic Acids Res.*, 2017.

[31] R. Schwanbeck, H. Xiao, and C. Wu, "Spatial Contacts and Nucleosome Step Movements Induced by the NURF Chromatin Remodeling Complex," *J. Biol. Chem.*, vol. 279, no. 38, pp. 39933–39941, Sep. 2004.

[32] A. Flaus and T. Owen-Hughes, "Dynamic properties of nucleosomes during thermal and ATP-driven mobilization.," *Mol. Cell. Biol.*, vol. 23, no. 21, pp. 7767–79, Nov. 2003.

[33] K. L. Jordan-Sciutto, J. M. Dragich, J. Caltagarone, D. J. Hall, and R. Bowser, "Fetal Alz-50 clone 1 (FAC1) protein interacts with the Myc-associated zinc finger protein (ZF87/MAZ) and alters its transcriptional activity.," *Biochemistry*, vol. 39, no. 12, pp. 3206–15, Mar. 2000.

[34] G. D. Strachan *et al.*, "Fetal Alz-50 clone 1 interacts with the human orthologue of the Kelch-like Ech-associated protein.," *Biochemistry*, vol. 43, no. 38, pp. 12113–22, Sep. 2004.

[35] P. Palma *et al.*, "Microarray Profiling of Mononuclear Peripheral Blood Cells Identifies Novel Candidate Genes Related to Chemoradiation Response in Rectal Cancer," *PLoS One*, vol. 8, no. 9, p. e74034, Sep. 2013.

[36] L. Richart *et al.*, "BPTF is required for c-MYC transcriptional activity and in vivo tumorigenesis," *Nat. Commun.*, vol. 7, p. 10153, Jan. 2016.

[37] D. G. MacArthur *et al.*, "A Systematic Survey of Loss-of-Function Variants in Human Protein-Coding Genes," *Science (80-.).*, vol. 335, no. 6070, pp. 823–828, Feb. 2012.

[38] Y. Buganim *et al.*, "A novel translocation breakpoint within the BPTF gene is associated with a pre-malignant phenotype," *PLoS One*, 2010.

[39] P. Stankiewicz *et al.*, "Haploinsufficiency of the Chromatin Remodeler BPTF Causes Syndromic Developmental and Speech Delay, Postnatal Microcephaly, and Dysmorphic Features," *Am. J. Hum. Genet.*, 2017.

[40] J. H. Lee, M. S. Kim, N. J. Yoo, and S. H. Lee, "BPTF, a chromatin remodeling-related gene, exhibits frameshift mutations in gastric and colorectal cancers ,"

APMIS, 2016.

[41] M. Roussy *et al.*, "NUP98-BPTF gene fusion identified in primary refractory acute megakaryoblastic leukemia of infancy," *Genes Chromosom. Cancer*, 2018.

[42] P. Badenhorst, M. Voas, I. Rebay, and C. Wu, "Biological functions of the ISWI chromatin remodeling complex NURF," *Genes Dev.*, 2002.

[43] J. Wysocka *et al.*, "A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling," *Nature*, vol. 442, no. 7098, pp. 86–90, May 2006.

[44] S. Y. Kwon, H. Xiao, C. Wu, and P. Badenhorst, "Alternative splicing of NURF301 generates distinct NURF chromatin remodeling complexes with altered modified histone binding specificities," *PLoS Genet.*, 2009.

[45] T. W. Harris *et al.*, "Wormbase: A comprehensive resource for nematode research," *Nucleic Acids Res.*, 2009.

[46] E. C. Andersen, X. Lu, and H. R. Horvitz, "C. elegans ISWI and NURF301 antagonize an Rb-like pathway in the determination of multiple cell fates.," *Development*, vol. 133, no. 14, pp. 2695–704, 2006.

[47] T. Blumenthal, "Trans-splicing and operons," *WormBook*, 2005.

[48] E. Guisbert, D. M. Czyz, K. Richter, P. D. McMullen, and R. I. Morimoto, "Identification of a Tissue-Selective Heat Shock Response Regulatory Network," *PLoS Genet.*, 2013.

[49] N. L. Bray, H. Pimentel, P. Melsted, and L. Pachter, "Near-optimal probabilistic RNA-seq quantification," *Nat. Biotechnol.*, 2016.

[50] J. Brunquell, S. Morris, Y. Lu, F. Cheng, and S. D. Westerheide, "The genome-wide role of HSF-1 in the regulation of gene expression in Caenorhabditis elegans," *BMC Genomics*, 2016.

[51] M. Uhlen *et al.*, "Building Predictive Models inRUsing thecaretPackage," *Nucleic Acids Res.*, 2011.

[52] X. Chen, Y. Shen, and R. E. Ellis, "Dependence of the sperm/oocyte decision on the nucleosome remodeling factor complex was acquired during recent caenorhabditis briggsae evolution," *Mol. Biol. Evol.*, 2014.

[53] D. J. Owen *et al.*, "The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase gcn5p.," *EMBO J.*, vol. 19, no. 22, pp. 6141–9, Nov. 2000.

[54] C. Frøkjær-Jensen *et al.*, "Random and targeted transgene insertion in Caenorhabditis elegans using a modified Mos1 transposon," *Nat. Methods*, vol. 11,

no. 5, pp. 529–534, May 2014.

[55] G. A. Tomlinson and M. Rothstein, “Nematode biochemistry. I. Culture methods,” *BBA - Biochim. Biophys. Acta*, 1962.

[56] A. Fodor, D. L. Riddle, F. Kenneth Nelson, and J. W. Golden, “Comparison of A New Wild-Type Caenorhabditis Briggsae With Laboratory Strains Of C. Briggsae And C. Elegans,” *Nematologica*, 1983.

[57] P. Zhang *et al.*, “A functional intronic variant in the tyrosine hydroxylase (TH) gene confers risk of essential hypertension in the Northern Chinese Han population,” *Clin. Sci.*, 2008.

[58] J. Naukkarinen *et al.*, “USF1 and dyslipidemias: Converging evidence for a functional intronic variant,” *Hum. Mol. Genet.*, 2005.

[59] K. Kiontke and W. Sudhaus, “Ecology of *Caenorhabditis* species.,” in *WormBook: The online review of* C. elegans *biology*, 2006.

[60] T. A. Manolio *et al.*, “Finding the missing heritability of complex diseases,” *Nature*. 2009.

[61] X. Li *et al.*, “Cats lack a sweet taste receptor.,” *J. Nutr.*, 2006.

[62] D. L. Stern and V. Orgogozo, “The loci of evolution: How predictable is genetic

evolution?," *Evolution*, vol. 62, no. 9. pp. 2155–2177, 2008.

[63] T. Goller, F. Vauti, S. Ramasamy, and H.-H. Arnold, "Transcriptional Regulator BPTF/FAC1 Is Essential for Trophoblast Differentiation during Early Mouse Development," *Mol. Cell. Biol.*, vol. 28, no. 22, pp. 6819–6827, Nov. 2008.